PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5:
C12N 15/51, C12Q 1/68
C12N 15/40, C12Q 1/70
A61K 39/29, C07K 13/00
G01N 33/576
(11) International Put (43) International Put (43) International Put (43) International Put (43)

(11) International Publication Number: WO 92/19743

(43) International Publication Date: 12 Novem

12 November 1992 (12.11.92)

(21) International Application Number:

PCT/US92/04036

(22) International Filing Date:

8 May 1992 (08.05.92)

(30) Priority data:

697,326

8 May 1991 (08.05.91)

US

(71) Applicant: CHIRON CORPORATION [US/US]; 4560 Horton Street, Emeryville, CA 94608 (US).

(72) Inventors: CHA, Tai-An; 964 Springview Circle, San Ramon, CA 94583 (US). BEALL, Eileen; 1150 Lincoln Avenue, #5, Walnut Creek, CA 94596 (US). IRVINE, Bruce; 3401 El Monte Drive, Concord, CA 94519 (US). KOLBERG, Janice; 131 Scots Valley, Hercules, CA 94547 (US). URDEA, Michael, S.; 100 Bunce Meadow Road, Alamo, CA 94501 (US).

(74) Agent: JANIUK, Anthony, J.; Wolf, Greenfield & Sacks, 600 Atlantic Avenue, Boston, MA 02210 (US).

(81) Designated States: AT (European patent), AU, BB, BE (European patent), BF (OAPI patent), BG, BJ (OAPI patent), BR, CA, CF (OAPI patent), CG (OAPI patent), CH (European patent), CI (OAPI patent), CM (OAPI patent), CS, DE (European patent), DK (European patent), ES (European patent), FI, FR (European patent), GA (OAPI patent), GB (European patent), GN (OAPI patent), GR (European patent), HU, IT (European patent), JP, KP, KR, LK, LU (European patent), MC (European patent), MG, ML (OAPI patent), MN, MR (OAPI patent), MW, NL (European patent), NO, PL, RO, RU, SD, SE (European patent), SN (OAPI patent), TD (OAPI patent), TG (OAPI patent).

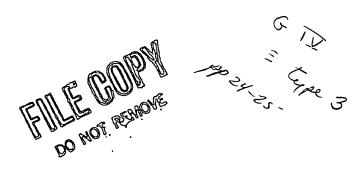
Published

Without international search report and to be republished upon receipt of that report.

(54) Title: HCV GENOMIC SEQUENCES FOR DIAGNOSTICS AND THERAPEUTICS

(57) Abstract

The present application features nucleic acid, peptide and antibody compositions relating to genotypes of hepatitis C virus and methods of using such compositions for diagnostic and therapeutic purposes.



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT Austria AU Australia BB Barbalon BE Belgium BF Burlana Fiso BF Greece HU Hungary Lialy CF Control African Republic CF Control African Republic CF Control African Republic CF Control Control CF Control	MC MI, MM MR MW NL NO PL RO RU SD SE SN SU TD TC US	Madagascar Mali Mongolia Manyitanla Malawi Nethertando Norway Poland Romania Russian Federation Sudan Sweden Senegal Soviet Union Chad Togo United States of America

HCV GENOMIC SEQUENCES FOR DIAGNOSTICS AND THERAPEUTICS

This application is a continuation-in-part of U.S. Serial No. 07/697,326 entitled "Polynucleotide Probes Useful for Screening for Hepatitis C Virus, filed May 8, 1991.

Technical Field

The invention relates to compositions and methods for the detection and treatment of hepatitis C virus, (HCV) infection, formerly referred to as blood-borne non-A, non-B hepatitis virus (NANBV) infection. More specifically, embodiments of the present invention feature compositions and methods for the detection of HCV, and for the development of vaccines for the prophylactic treatment of infections of HCV, and development of antibody products for conveying passive immunity to HCV.

20

25

Background of the Invention

The prototype isolate of HCV was characterized in U.S. Patent Application Serial No. 122,714 (See also EPO Publication No. 318,216). As used herein, the term "HCV" includes new isolates of the same viral species. The term "HCV-1" referred to in U.S. Patent Application Serial No. 122,714.

15

20

25

- 2 -

HCV is a transmissible disease distinguishable from other forms of viral-associated liver diseases, including that caused by the known hepatitis viruses, i.e., hepatitis A virus (HAV), hepatitis B virus (HBV), and delta hepatitis virus (HDV), as well as the hepatitis induced by cytomegalovirus (CMV) or Epstein-Barr virus (EBV). HCV was first identified in individuals who had received blood transfusions.

The demand for sensitive, specific methods for screening and identifying carriers of HCV and HCV contaminated blood or blood products is significant. Post-transfusion hepatitis (PTH) occurs in approximately 10% of transfused patients, and HCV accounts for up to 90% of these cases. The disease frequently progresses to chronic liver damage (25-55%).

Patient care as well as the prevention of transmission of HCV by blood and blood products or by close personal contact require reliable screening, diagnostic and prognostic tools to detect nucleic acids, antigens and antibodies related to HCV.

Information in this application suggests the HCV has several genotypes. That is, the genetic information of the HCV virus may not be totally identical for all HCV, but encompasses groups with differing genetic information.

Genetic information is stored in thread-like molecules of DNA and RNA. DNA consists of covalently

linked chains of deoxyribonucleotides and RNA consists of covalently linked chains of ribonucleotides. Each nucleotide is characterized by one of four bases: adenine (A), quanine (G), thymine (T), and cytosine 5 (C). The bases are complementary in the sense that, due to the orientation of functional groups, certain base pairs attract and bond to each other through hydrogen bonding and w-stacking interactions. Adenine in one strand of DNA pairs with thymine in an opposing complementary strand. Guanine in one strand 10 of DNA pairs with cytosine in an opposing complementary strand. In RNA, the thymine base is replaced by uracil (U) which pairs with adenine in an opposing complementary strand. The genetic code of living organism is carried in the sequence of base pairs. 15 Living cells interpret, transcribe and translate the information of nucleic acid to make proteins and peptides.

The HCV genome is comprised of a single positive strand of RNA. The HCV genome possesses a continuous, translational open reading frame (ORF) that encodes a polyprotein of about 3,000 amino acids. In the ORF, the structural protein(s) appear to be encoded in approximately the first quarter of the N-terminus region, with the majority of the polyprotein 25 responsible for non-structural proteins.

10

15

20

_ 4 -

The HCV polyprotein comprises, from the amino terminus to the carboxy terminus, the nucleocapsid protein (C), the envelope protein (E), and the non-structural proteins (NS) 1, 2 (b), 3, 4 (b), and 5.

HCV of differing genotypes may encode for proteins which present an altered response to host immune systems. HCV of differing genotypes may be difficult to detect by immuno diagnostic techniques and nucleic acid probe techniques which are not specifically directed to such genotype.

Definitions for selected terms used in the application are set forth below to facilitate an understanding of the invention. The term "corresponding" means homologous to or complementary to a particular sequence of nucleic acid. As between nucleic acids and peptides, corresponding refers to amino acids of a peptide in an order derived from the sequence of a nucleic acid or its complement.

The term "non-naturally occurring nucleic acid"
refers to a portion of genomic nucleic acid, cDNA,
semisynthetic nucleic acid, or synthetic origin nucleic
acid which, by virtue of its origin or manipulation:
(1) is not associated with all of a nucleic acid with
which it is associated in nature, (2) is linked to a
nucleic acid or other chemical agent other than that to

which it is linked in nature, or (3) does not occur in nature.

Similarly the term, "a non-naturally occurring peptide" refers to a portion of a large naturally occurring peptide or protein, or semi-synthetic or synthetic peptide, which by virtue of its origin or manipulation (1) is not associated with all of a peptide with which it is associated in nature, (2) is linked to peptides, functional groups or chemical agents other than that to which it is linked in nature, or (3) does not occur in nature.

The term "primer" refers to a nucleic acid which is capable of initiating the synthesis of a larger nucleic acid when placed under appropriate conditions.

The primer will be completely or substantially complementary to a region of the nucleic acid to be copied. Thus, under conditions conducive to hybridization, the primer will anneal to a complementary region of a larger nucleic acid. Upon addition of suitable reactants, the primer is extended by the polymerizing agent to form a copy of the larger nucleic acid.

The term "binding pair" refers to any pair of molecules which exhibit mutual affinity or binding
25 capacity. For the purposes of the present application, the term "ligand" will refer to one molecule of the binding pair, and the term "antiligand" or "receptor"

15

20

25

or "target" will refer to the opposite molecule of the binding pair. For example, with respect to nucleic acids, a binding pair may comprise two complementary nucleic acids. One of the nucleic acids may be designated the ligand and the other strand is designated the antiligand receptor or target. The designation of ligand or antiligand is a matter of arbitrary convenience. Other binding pairs comprise, by way of example, antigens and antibodies, drugs and drug receptor sites and enzymes and enzyme substrates, to name a few.

The term "label" refers to a molecular moiety capable of detection including, by way of example, without limitation, radioactive isotopes, enzymes, luminescent agents, precipitating agents, and dyes.

The term "support" includes conventional supports such as filters and membranes as well as retrievable supports which can be substantially dispersed within a medium and removed or separated from the medium by immobilization, filtering, partitioning, or the like. The term "support means" refers to supports capable of being associated to nucleic acids, peptides or antibodies by binding partners, or covalent or noncovalent linkages.

A number of HCV strains and isolates have been identified. When compared with the sequence of the original isolate derived from the USA ("HCV-1"; see

Q.-L. Choo et al. (1989) Science 244:359-362, Q.-L. Choo et al. (1990) Brit. Med. Bull. 46:423-441, Q.-L. Choo et al., Proc. Natl. Acad. Sci. 88:2451-2455 (1991), and E.P.O. Patent Publication No. 318,216, cited supra), it was found that a Japanese isolate ("HCV J1") differed significantly in both nucleotide and polypeptide sequence within the NS3 and NS4 regions. This conclusion was later extended to the NS5 and envelope (E1/S and E2/NS1) regions (see K. Takeuchi 10 et al., J. Gen. Virol. (1990) 71:3027-3033, Y. Kubo, Mucl. Acids. Res. (1989) 17:10367-10372, and K. Takeuchi et al., Gene (1990) 91:287-291). The former group of isolates, originally identified in the United States, is termed "Genotype I" throughout the present disclosure, while the latter group of isolates, 15 initially identified in Japan, is termed "Genotype II" herein.

Brief Description of the Invention

The present invention features compositions of matter comprising nucleic acids and peptides corresponding to the HCV viral genome which define different genotypes. The present invention also features methods of using the compositions corresponding to sequences of the HCV viral genome which define different genotypes described herein.

10

15

20

25

-8-

A. Nucleic acid compositions

The nucleic acid of the present invention, corresponding to the HCV viral genome which define different genotypes, have utility as probes in nucleic acid hybridization assays, as primers for reactions involving the synthesis of nucleic acid, as binding partners for separating HCV viral nucleic acid from other constituents which may be present, and as anti-sense nucleic acid for preventing the transcription or translation of viral nucleic acid.

One embodiment of the present invention features a composition comprising a non-naturally occurring nucleic acid having a nucleic acid sequence of at least eight nucleotides corresponding to a non-HCV-1 nucleotide sequence of the hepatitis C viral genome. Preferably, the nucleotide sequence is selected from a sequence present in at least one region consisting of the NS5 region, envelope 1 region, 5'UT region, and the core region.

Preferably, with respect to sequences which correspond to the NS5 region, the sequence is selected from a sequence within a sequence numbered 2-22. The sequence numbered 1 corresponds to HCV-1. Sequences numbered 1-22 are defined in the Sequence Listing of the application.

Preferably, with respect to sequences corresponding to the envelope 1 region, the sequence is

15

selected from a sequence within sequences numbered 24-32. Sequence No. 23 corresponds to HCV-1. Sequences numbered 23-32 are set forth in the Sequence Listing of the application.

Preferably, with respect to the sequences which correspond to the 5'UT regions, the sequence is selected from a sequence within sequences numbered 34-51. Sequence No. 33 corresponds to HCV-1. Sequence No. 33-51 are set forth in the Sequence Listing of this application.

Preferably, with respect to the sequences which correspond to the core region, the sequence is selected from a sequence within the sequences numbered 53-66. Sequence No. 52 corresponds to HCV-1. Sequences 52-66 are set forth in the Sequence Listing of this application.

The compositions of the present invention form hybridization products with nucleic acid corresponding to different genotypes of HCV.

20 HCV has at least five genotypes, which will be referred to in this application by the designations GI-GV. The first genotype, GI, is exemplified by sequences numbered 1-6, 23-25, 33-38 and 52-57. The second genotype, GII, is exemplified by the sequences numbered 7-12, 26-28, 39-45 and 58-64. The third genotype, GIII, is exemplified by sequences numbered 13-17, 32, 46-47 and 65-66. The fourth genotype, GIV,

is exemplified by sequences numbered 20-22, and 29-31 and 48-49. The fifth genotype, GV, is exemplified by sequences numbered 18, 19, 50 and 51.

One embodiment of the present invention features compositions comprising a nucleic acid having a sequence corresponding to one or more sequences which exemplify a genotype of HCV.

Method of forming a Hybridization Product Embodiments of the present invention also feature 10 a method of forming a hybridization product with nucleic acid having a sequence corresponding to HCV nucleic acid. One method comprises the steps of placing a non-naturally occurring nucleic acid having a non-HCV-1 sequence corresponding to HCV nucleic acid 15 under conditions in which hybridization may occur. The non-naturally occurring nucleic acid is capable of forming a hybridization product with HCV nucleic acid, under hybridization conditions. The method further comprises the step of imposing hybridization conditions 20 to form a hybridization product in the presence of nucleic acid corresponding to a region of the HCV genome.

The formation of a hybridization product has utility for detecting the presence of one or more genotypes of HCV. Preferably, the non-naturally occurring nucleic acid forms a hybridization product

15

with nucleic acid of HCV in one or more regions comprising the NS5 region, envelope 1 region, 5'UT region and the core region. To detect the hybridization product, it is useful to associate the non-naturally occurring nucleic acid with a label. The formation of the hybridization product is detected by separating the hybridization product from labeled non-naturally occurring nucleic acid, which has not formed a hybridization product.

The formation of a hybridization product has utility as a means of separating one or more genotypes of HCV nucleic acid from other constituents potentially present. For such applications, it is useful to associate the non-naturally occurring nucleic acid with a support for separating the resultant hybridization product from the the other constituents.

Nucleic acid "sandwich assays" employ one nucleic acid associated with a label and a second nucleic acid associated with a support. An embodiment of the present invention features a sandwich assay comprising two nucleic acids, both have sequences which correspond to HCV nucleic acids; however, at least one non-naturally occurring nucleic acid has a sequence corresponding to non-HCV-1 HCV nucleic acid. At least one nucleic acid is capable of associating with a label, and the other is capable of associating with a support. The support associated non-naturally

10

15

20

25

occurring nucleic acid is used to separate the hybridization products which include an HCV nucleic acid and the non-naturally occurring nucleic acid having a non-HCV-1 sequence.

One embodiment of the present invention features a method of detecting one or more genotypes of HCV. method comprises the steps of placing a non-naturally occurring nucleic acid under conditions which hybridization may occur. The non-naturally occurring nucleic acid is capable of forming a hybridization product with nucleic acid from one or more genotypes of HCV. The first genotype, GI, is exemplified by sequences numbered 1-6, 23-25, 33-38 and 52-57. The second genotype, GII, is exemplified by the sequences numbered 7-12, 26-28, 39-45 and 58-64. The third genotype, GIII, is exemplified by sequences numbered 13-17, 32, 46-47 and 65-66. The fourth genotype, GIV, is exemplified sequences numbered 20-22 and 29-31. fifth genotype, GV, is exemplified by sequences numbered 18, 19, 50 and 51.

The hybridization product of HCV nucleic acid with a non-naturally occurring nucleic acid having non-HCV-1 sequence corresponding to sequences within the HCV genome has utility for priming a reaction for the synthesis of nucleic acid.

The hybridization product of HCV nucleic acid with a non-naturally occurring nucleic acid having a

sequence corresponding to a particular genotype of HCV has utility for priming a reaction for the synthesis of nucleic acid of such genotype. In one embodiment, the synthesized nucleic acid is indicative of the presence of one or more genotypes of HCV.

The synthesis of nucleic acid may also facilitate cloning of the nucleic acid into expression vectors which synthesize viral proteins.

anti-sense agents for preventing the transcription or translation of viral nucleic acid. The formation of a hybridization product of a non-naturally occurring nucleic acid having sequences which correspond to a particular genotype of HCV genomic sequencing with HCV nucleic acid may block translation or transcription of such genotype. Therapeutic agents can be engineered to include all five genotypes for inclusivity.

C. Peptide and antibody composition

A further embodiment of the present invention

20 features a composition of matter comprising a
non-naturally occurring peptide of three or more amino
acids corresponding to a nucleic acid having a
non-HCV-1 sequence. Preferably, the non-HCV-1 sequence
corresponds with a sequence within one or more regions

25 consisting of the NS5 region, the envelope 1 region,
the 5'UT region, and the core region.

15

20

25

Preferably, with respect to peptides corresponding to a nucleic acid having a non-HCV-1 sequence of the NS5 region, the sequence is within sequences numbered 2-22. The sequence numbered 1 corresponds to HCV-1. Sequences numbered 1-22 are set forth in the Sequence Listing.

Preferably, with respect to peptides corresponding to a nucleic acid having a non-HCV-1 sequence of the envelope 1 region, the sequence is within sequences numbered 24-32. The sequence numbered 23 corresponds to HCV-1. Sequences numbered 23-32 are set forth in the Sequence Listing.

Preferably, with respect to peptides corresponding to a nucleic acid having a non-HCV-1 sequence directed to the core region, the sequence is within sequences numbered 53-66. Sequence numbered 52 corresponds to HCV-1. Sequences numbered 52-66 are set forth in the Sequence Listing.

The further embodiment of the present invention features peptide compositions corresponding to nucleic acid sequences of a genotype of HCV. The first genotype, GI, is exemplified by sequences numbered 1-6, 23-25, 33-38 and 52-57. The second genotype, GII, is exemplified by the sequences numbered 7-12, 26-28, 39-45 and 58-64. The third genotype, GIII, is exemplified by sequences numbered 13-17, 32, 46-47 and 65-66. The fourth genotype, GIV, is exemplified

15

20

25

sequences numbered 20-22, 29-31, 48 and 49. The fifth genotype, GV, is exemplified by sequences numbered 18, 19, 50 and 51.

The non-naturally occurring peptides of the present invention are useful as a component of a vaccine. The sequence information of the present invention permits the design of vaccines which are inclusive for all or some of the different genotypes of HCV. Directing a vaccine to a particular genotype allows prophylactic treatment to be tailored to maximize the protection to those agents likely to be encountered. Directing a vaccine to more than one genotype allows the vaccine to be more inclusive.

The peptide compositions are also useful for the development of specific antibodies to the HCV proteins. One embodiment of the present invention features as a composition of matter, an antibody to peptides corresponding to a non-HCV-1 sequence of the HCV genome. Preferably, the non-HCV-1 sequence is selected from the sequence within a region consisting of the NS5 region, the envelope 1 region, and the core region. There are no peptides associated with the untranslated 5'UT region.

Preferably, with respect to antibodies directed to peptides of the NS5 region, the peptide corresponds to a sequence within sequences numbered 2-22. Preferably, with respect to antibodies directed to a peptide

15

20

25

corresponding to the envelope 1 region, the peptide corresponds to a sequence within sequences numbered 24-32. Preferably, with respect to the antibodies directed to peptides corresponding to the core region, the peptide corresponds to a sequence within sequences numbered 53-66.

Antibodies directed to peptides which reflect a particular genotype have utility for the detection of such genotypes of HCV and therapeutic agents.

One embodiment of the present invention features an antibody directed to a peptide corresponding to nucleic acid having sequences of a particular genotype. The first genotype, GI, is exemplified by sequences numbered 1-6, 23-25, 33-38 and 52-57. The second genotype, GII, is exemplified by the sequences numbered 7-12, 26-28, 39-45 and 58-64. The third genotype, GIII, is exemplified by sequences numbered 13-17, 32, 46-47 and 65-66. The fourth genotype, GIV, is exemplified sequences numbered 20-22, 29-31, 48 and 49. The fifth genotype, GV, is exemplified by sequences numbered 18, 19, 50 and 51.

Individuals skilled in the art will readily recognize that the compositions of the present invention can be packaged with instructions for use in the form of a kit for performing nucleic acid hybridizations or immunochemical reactions.

The present invention is further described in the following figures which illustrate sequences demonstrating genotypes of HCV. The sequences are designated by numerals 1-145, which numerals and sequences are consistent with the numerals and sequences set forth in the Sequence Listing. Sequences 146 and 147 facilitate the discussion of an assay which numerals and sequences are consistent with the numerals and sequences set forth in the Sequence Listing.

10

25

Brief Description of the Figures and Sequence Listing
Figure 1 depicts schematically the genetic
organization of HCV;

Figure 2 sets forth nucleic acid sequences

15 numbered 1-22 which sequences are derived from the NS5
region of the HCV viral genome;

Figure 3 sets forth nucleic acid sequences numbered 23-32 which sequences are derived from the envelope 1 region of the HCV viral genome;

20 Figure 4 sets forth nucleic acid sequences numbered 33-51 which sequences are derived from the 5'UT region of the HCV viral genome; and,

Pigure 5 sets forth nucleic acid sequences numbered 52-66 which sequences are derived from the core region of the HCV viral genome.

The Sequence Listing sets forth the sequences of sequences numbered 1-147.

- 18 -

Detailed Description of the Invention

The present invention will be described in detail as as nucleic acid having sequences corresponding to the HCV genome and related peptides and binding partners, for diagnostic and therapeutic applications.

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of chemistry, molecular biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature. See e.g., Maniatis, Fitsch & Sambrook, Molecular Cloning; A Laboratory Manual (1982); DNA Cloning, Volumes I and II (D.N Glover ed. 1985); Oligonucleotide Synthesis (M.J. Gait ed, 1984); Nucleic Acid Hybridization (B.D. Hames & S.J. Higgins eds. 15 1984); the series, Methods in Enzymology (Academic Press, Inc.), particularly Vol. 154 and Vol. 155 (Wu and Grossman, eds.).

The cDNA libraries are derived from nucleic acid sequences present in the plasma of an HCV-infected 20 chimpanzee. The construction of one of these libraries, the "c" library (ATCC No. 40394), is described in PCT Pub. No. WO90/14436. The sequences of the library relevant to the present invention are set forth herein as sequence numbers 1, 23, 33 and 52. 25

Nucleic acids isolated or synthesized in accordance with features of the present invention are

10

15

- 20

25

useful, by way of example without limitation as probes, primers, anti-sense genes and for developing expression systems for the synthesis of peptides corresponding to such sequences.

The nucleic acid sequences described define genotypes of HCV with respect to four regions of the viral genome. Figure 1 depicts schematically the organization of HCV. The four regions of particular interest are the NS5 region, the envelope 1 region, the 5'UT region and the core region.

The sequences set forth in the present application as sequences numbered 1-22 suggest at least five genotypes in the NS5 region. Sequences numbered 1-22 are depicted in Figure 2 as well as the Sequence Listing. Each sequence numbered 1-22 is derived from nucleic acid having 340 nucleotides from the NS5 region.

The five genotypes are defined by groupings of the sequences defined by sequence numbered 1-22. For convenience, in the present application, the different genotypes will be assigned roman numerals and the letter "G".

The first genotype (GI) is exemplified by sequences within sequences numbered 1-6. A second genotype (GII) is exemplified by sequences within sequences numbered 7-12. A third genotype (GIII) is exemplified by the sequences within sequences numbered 13-17. A fourth genotype (GIV) is exemplified by

15

sequences within sequences numbered 20-22. A fifth genotype (GV) is exemplified by sequences within sequences numbered 18 and 19.

The sequences set forth in the present application as sequences numbered 23-32 suggest at least four genotypes in the envelope 1 region of HCV. Sequences numbered 23-32 are depicted in Figure 3 as well as in the Sequence Listing. Each sequence numbered 23-32 is derived from nucleic acid having 100 nucleotides from the envelope 1 region.

A first envelope 1 genotype group (GI) is exemplified by the sequences within the sequences numbered 23-25. A second envelope 1 genotype (GII) region is exemplified by sequences within sequences numbered 26-28. A third envelope 1 genotype (GIII) is exemplified by the sequences within sequences numbered 32. A fourth envelope 1 genotype (GIV) is exemplified by the sequences within sequence numbered 29-31.

The sequences set forth in the present application
as sequences numbered 33-51 suggest at least three
genotypes in the 5'UT region of HCV. Sequences
numbered 33-51 are depicted in Figure 4 as well as in
the Sequence Listing. Each sequence numbered 33-51 is
derived from the nucleic acid having 252 nucleotides
from the 5'UT region, although sequences 50 and 51 are
somewhat shorter at approximately 180 nucleotides.

15

20

25

The first 5'UT genotype (GI) is exemplified by the sequences within sequences numbered 33-38. A second 5'UT genotype (GII) is exemplified by the sequences within sequences numbered 39-45. A third 5'UT genotype (GIII) is exemplified by the sequences within sequences numbered 46-47. A fourth 5'UT genotype (GIV) is exemplified by sequences within sequences humbered 48 and 49. A fifth 5'UT genotype (GV) is exemplified by sequences within sequences numbered 50 and 51.

The sequences numbered 48-62 suggest at least three genotypes in the core region of HCV. The sequences numbered 52-66 are depicted in Figure 5 as well as in the Sequence Listing.

The first core region genotype (GI) is exemplified by the sequences within sequences numbered 52-57. The second core region genotype (GII) is exemplified by sequences within sequences numbered 58-64. The third core region genotype (GIII) is exemplified by sequences within sequences numbered 65 and 66. Sequences numbered 52-65 are comprised of 549 nucleotides. Sequence numbered 66 is comprised of 510 nucleotides.

The various genotypes described with respect to each region are consistent. That is, HCV having features of the first genotype with respect to the NS5 region will substantially conform to features of the first genotype of the envelope 1 region, the 5'UT region and the core region.

Nucleic acid isolated or synthesized in accordance with the sequences set forth in sequence numbers 1-66 are useful as probes, primers, capture ligands and anti-sense agents. As probes, primers, capture ligands and anti-sense agents, the nucleic acid wil normally comprise approximately eight or more nucleotides for specificity as well as the ability to form stable hybridization products.

10 Probes

15

20

A nucleic acid isolated or synthesized in accordance with a sequence defining a particular genotype of a region of the HCV genome can be used as a probe to detect such genotype or used in combination with other nucleic acid probes to detect substantially all genotypes of HCV.

With the sequence information set forth in the present application, sequences of eight or more nucleotides are identified which provide the desired inclusivity and exclusivity with respect to various genotypes within HCV, and extraneous nucleic acid sequences likely to be encountered during hybridization conditions.

Individuals skilled in the art will readily
recognize that the nucleic acid sequences, for use as
probes, can be provided with a label to facilitate
detection of a hybridization product.

Capture Ligand

For use as a capture ligand, the nucleic acid selected in the manner described above with respect to probes, can be readily associated with supports. The manner in which nucleic acid is associated with supports is well known. Nucleic acid having sequences corresponding to a sequence within sequences numbered 1-66 have utility to separate viral nucleic acid of one genotype from the nucleic acid of HCV of a different genotype. Nucleic acid isolated or synthesized in accordance with sequences within sequences numbered 1-66, used in combinations, have utility to capture substantially all nucleic acid of all HCV genotypes.

15 Primers

10

Nucleic acid isolated or synthesized in accordance with the sequences described herein have utility as primers for the amplification of HCV sequences. With respect to polymerase chain reaction (PCR) techniques, nucleic acid sequences of eight or more nucleotides corresponding to one or more sequences of sequences numbered 1-66 have utility in conjunction with suitable enzymes and reagents to create copies of the viral nucleic acid. A plurality of primers having different sequences corresponding to more than one genotype can be used to create copies of viral nucleic acid for such genotypes.

The copies can be used in diagnostic assays to detect HCV virus. The copies can also be incorporated into cloning and expression vectors to generate polypeptides corresponding to the nucleic acid synthesized by PCR, as will be described in greater detail below.

Anti-sense

Nucleic acid isolated or synthesized in accordance with the sequences described herein have utility as anti-sense genes to prevent the expression of HCV.

Nucleic acid corresponding to a genotype of HCV is loaded into a suitable carrier such as a liposome for introduction into a cell infected with HCV. A nucleic acid having eight or more nucleotides is capable of binding to viral nucleic acid or viral messenger RNA. Preferably, the anti-sense nucleic acid is comprised of 30 or more nucleotides to provide necessary stability of a hybridization product of viral nucleic acid or viral messenger RNA. Methods for loading anti-sense nucleic acid is known in the art as exemplified by U.S. Patent 4,241,046 issued December 23, 1980 to Papahadjopoulos et al.

25 Peptide Synthesis

Nucleic acid isolated or synthesized in accordance with the sequences described herein have utility to

generate peptides. The sequences exemplified by sequences numbered 1-32 and 52-66 can be cloned into suitable vectors or used to isolate nucleic acid. The isolated nucleic acid is combined with suitable DNA linkers and cloned into a suitable vector. The vector can be used to transform a suitable host organism such as <u>E. coli</u> and the peptide encoded by the sequences isolated.

Molecular cloning techniques are described in the text Molecular Cloning: A Laboratory Manual, Maniatis et al., Coldspring Harbor Laboratory (1982).

The isolated peptide has utility as an antigenic substance for the development of vaccines and antibodies directed to the particular genotype of HCV.

15

10

Vaccines and Antibodies

The peptide materials of the present invention have utility for the development of antibodies and vaccines.

The availability of cDNA sequences, or nucleotide sequences derived therefrom (including segments and modifications of the sequence), permits the construction of expression vectors encoding antigenically active regions of the peptide encoded in either strand. The antigenically active regions may be derived from the NS5 region, envelope 1 regions, and the core region.

15

20

Fragments encoding the desired peptides are derived from the cDNA clones using conventional restriction digestion or by synthetic methods, and are ligated into vectors which may, for example, contain portions of fusion sequences such as beta galactosidase or superoxide dismutase (SOD), preferably SOD. Methods and vectors which are useful for the production of polypeptides which contain fusion sequences of SOD are described in European Patent Office Publication number 0196056, published October 1, 1986.

Any desired portion of the HCV cDNA containing an open reading frame, in either sense strand, can be obtained as a recombinant peptide, such as a mature or fusion protein; alternatively, a peptide encoded in the cDNA can be provided by chemical synthesis.

The DNA encoding the desired peptide, whether in fused or mature form, and whether or not containing a signal sequence to permit secretion, may be ligated into expression vectors suitable for any convenient host. Both sukaryotic and prokaryotic host systems are presently used in forming recombinant peptides. The peptide is then isolated from lysed cells or from the culture medium and purified to the extent needed for its intended use. Purification may be by techniques known in the art, for example, differential extraction, salt fractionation, chromatography on ion exchange resins, affinity chromatography, centrifugation, and

the like. See, for example, Methods in Enzymology for a variety of methods for purifying proteins. Such peptides can be used as diagnostics, or those which give rise to neutralizing antibodies may be formulated into vaccines. Antibodies raised against these peptides can also be used as diagnostics, or for passive immunotherapy or for isolating and identifying HCV.

An antigenic region of a peptide is generally relatively small--typically 8 to 10 amino acids or less 10 in length. Pragments of as few as 5 amino acids may characterize an antigenic region. These segments may correspond to NS5 region, envelope 1 region, and the core region of the HCV genome. The 5'UT region is not known to be translated. Accordingly, using the cDNAs 15 of such regions, DNAs encoding short segments of HCV peptides corresponding to such regions can be expressed recombinantly either as fusion proteins, or as isolated peptides. In addition, short amino acid sequences can be conveniently obtained by chemical synthesis. In 20 instances wherein the synthesized peptide is correctly configured so as to provide the correct epitope, but is too small to be immunogenic, the peptide may be linked to a suitable carrier.

25 A number of techniques for obtaining such linkage are known in the art, including the formation of disulfide linkages using N-succinimidyl-3-(2-

pyridylthio)propionate (SPDP) and succinimidyl . 4-(N-maleimido-methyl)cyclohexane-l-carboxylate (SMCC) obtained from Pierce Company, Rockford, Illinois, (if the peptide lacks a sulfhydryl group, this can be provided by addition of a cysteine residue). These reagents create a disulfide linkage between themselves and peptide cysteine residues on one protein and an amide linkage through the epsilon-amino on a lysine, or other free amino group in the other. 'A variety of such disulfide/amide-forming agents are known. See, for 10 example, Immun Rev (1982) 62:185. Other bifunctional coupling agents form a thioether rather than a disulfide linkage. Many of these thio-ether-forming agents are commercially available and include reactive esters of 6-maleimidocaprioc acid, 2-bromoacetic acid, 15 2-iodoacetic acid, 4-N-maleimido-methyl)cyclohexane-1carboxylic acid, and the like. The carboxyl groups can be activated by combining them with succinimide or 1-hydroxyl-2 nitro-4-sulfonic acid, sodium salt. Additional methods of coupling antigens employs the 20 rotavirus/"binding peptide" system described in EPO Pub. No. 259,149, the disclosure of which is incorporated herein by reference. The foregoing list is not meant to be exhaustive, and modifications of the named compounds can clearly be used. 25

Any carrier may be used which does not itself induce the production of antibodies harmful to the

host. Suitable carriers are typically large, slowly metabolized macromolecules such as proteins; polysaccharides, such as latex functionalized Sepharose, agarose, cellulose, cellulose beads and the like; polymeric amino acids, such as polyglutamic acid, polylysine, and the like; amino acid copolymers; and inactive virus particles. Especially useful protein substrates are serum albumins, keyhole limpet hemocyanin, immunoglobulin molecules, thyroglobulin, ovalbumin, tetanus toxoid, and other proteins well known to those skilled in the art.

Peptides comprising HCV amino acid sequences encoding at least one viral epitope derived from the NS5, envelope 1, and core region are useful 15 immunological reagents. The 5'UT region is not known to be translated. For example, peptides comprising such truncated sequences can be used as reagents in an immunoassay. These peptides also are candidate subunit antigens in compositions for antiserum production or vaccines. While the truncated sequences can be 20 produced by various known treatments of native viral protein, it is generally preferred to make synthetic or recombinant peptides comprising HCV sequence. Peptides comprising these truncated HCV sequences can be made up entirely of HCV sequences (one or more epitopes, either 25 contiguous or noncontiguous), or HCV sequences and heterologous sequences in a fusion protein. Useful

heterologous sequences include sequences that provide for secretion from a recombinant host, enhance the immunological reactivity of the HCV epitope(s), or facilitate the coupling of the polypeptide to an immunoassay support or a vaccine carrier. See, E.G., EPO Pub. No. 116,201; U.S. Pat. No. 4,722,840; EPO Pub. No. 259,149; U.S. Pat. No. 4,629,783.

The size of peptides comprising the truncated HCV sequences can vary widely, the minimum size being a sequence of sufficient size to provide an HCV epitope, 10 while the maximum size is not critical. For convenience, the maximum size usually is not substantially greater than that required to provide the desired HCV epitopes and function(s) of the heterologous sequence, if any. Typically, the truncated HCV amino acid sequence will range from about 5 to about 100 amino acids in length. More typically, however, the HCV sequence will be a maximum of about 50 amino acids in length, preferably a maximum of about 30 amino acids. It is usually desirable to select HCV sequences of at least about 10, 12 or 15 amino acids, up to a maximum of about 20 or 25 amino acids.

HCV amino acid sequences comprising epitopes can be identified in a number of ways. For example, the entire protein sequence corresponding to each of the NSS, envelope 1, and core regions can be screened by preparing a series of short peptides that together span

the entire protein sequence of such regions. By starting with, for example, peptides of approximately 100 amino acids, it would be routine to test each peptide for the presence of epitope(s) showing a desired reactivity, and then testing progressively smaller and overlapping fragments from an identified peptides of 100 amino acids to map the epitope of interest. Screening such peptides in an immunoassay is within the skill of the art. It is also known to carry out a computer analysis of a protein sequence to identify potential epitopes, and then prepare peptides comprising the identified regions for screening.

The immunogenicity of the epitopes of HCV may also be enhanced by preparing them in mammalian or yeast 15 systems fused with or assembled with particle-forming proteins such as, for example, that associated with hepatitis B surface antigen. See, e.g., US 4,722,840. Constructs wherein the HCV epitope is linked directly to the particle-forming protein coding sequences produce hybrids which are immunogenic with respect to the HCV epitope. In addition, all of the vectors prepared include epitopes specific to HBV, having various degrees of immunogenicity, such as, for example, the pre-S peptide. Thus, particles constructed from particle forming protein which include HCV sequences are immunogenic with respect to HCV and HBV.

Hepatitis surface antigen (HBSAg) has been shown to be formed and assembled into particles in S: cerevisiae (P. Valenzuela et al. (1982)), as well as in, for example, mammalian cells (P. Valenzuela et al. 1984)). The formation of such particles has been shown to enhance the immunogenicity of the monomer subunit. The constructs may also include the immunodominant epitope of HBSAg, comprising the 55 amino acids of the presurface (pre-S) region. Neurath et al. (1984). Constructs of the pre-S-HBSAg particle expressible in 10 yeast are disclosed in EPO 174,444, published March 19, 1986; hybrids including heterologous viral sequences for yeast expression are disclosed in EPO 175,261, published March 26, 1966. These constructs may also be expressed in mammalian cells such as Chinese hamster 15 ovary (CHO) cells using an SV40-dihydrofolate reductase vector (Michelle et al. (1984)).

In addition, portions of the particle-forming protein coding sequence may be replaced with codons encoding an HCV epitope. In this replacement, regions which are not required to mediate the aggregation of the units to form immunogenic particles in yeast of mammals can be deleted, thus eliminating additional HEV antigenic sites from competition with the HCV epitope.

25

20

Vaccines

Vaccines may be prepared from one or more

10

- 33 -

immunogenic peptides derived from HCV. The observed homology between HCV and Flaviviruses provides information concerning the peptides which are likely to be most effective as vaccines, as well as the regions of the genome in which they are encoded.

Multivalent vaccines against HCV may be comprised of one or more epitopes from one or more proteins derived from the NS5, envelope 1, and core regions. In particular, vaccines are contemplated comprising one or more HCV proteins or subunit antigens derived from the NS5, envelope 1, and core regions. The 5'UT region is not known to be translated.

The preparation of vaccines which contain an immunogenic peptide as an active ingredient, is known to one skilled in the art. Typically, such vaccines 15 are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid prior to injection may also be prepared. The preparation may also be emulsified, or the protein encapsulated in liposomes. The active 20 immunogenic ingredients are often mixed with excipients which are pharmaceutically acceptable and compatible with the active ingredient. Suitable excipients are, for example, water, saline, dextrose, glycerol, ethanol, or the like and combinations thereof. In 25 addition, if desired, the vaccine may contain minor amounts of auxiliary substances such as wetting or

25

emulsifying agents, pH buffering agents, and/or adjuvants which enhance the effectiveness of the vaccine. Examples of adjuvants which may be effective include but are not limited to: aluminum hydroxide, N-acetyl-muramyl-L-theronyl-D- isoglutamine (thr-MDP). N-acetyl-nor-muramyl-L-alanyl- D-isoglutamine (CGP 11637, referred to as nor-MDP), N- acetylmuramyl-Lalanyl-D-isoglutaminyl-L-alanine-2-(1- 2-dipalmitoyl -sn-glycero-3-hydroxyphosphoryloxy)- ethylamine (CGP 19835A, referred to as MTP-PE), and RIBI, which 10 contains three components extracted from bacteria, monophosphoryl lipid A, trehalose dimycolate and cell wall skeleton (MPL+TDM+CWS) in a 2% squalene/Tween 80 emulsion. The effectiveness of an adjuvant may be determined by measuring the amount of antibodies 15 directed against an immunogenic peptide containing an HCV antigenic sequence resulting from administration of this peptide in vaccines which are also comprised of the various adjuvants.

The vaccines are conventionally administered parenterally, by injection, for example, either subcutaneously or intramuscularly. Additional formulations which are suitable for other modes of administration include suppositories and, in some cases, oral formulations. For suppositories, traditional binders and carriers may include, for example, polyalkylene glycols or triglycerides; such

suppositories may be formed from mixtures containing the active ingredient in the range of 0/5% to 10%, preferably 1%-2%. Oral formulations include such normally employed excipients as, for example,

pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, sodium saccharine, cellulose, magnesium carbonate, and the like.

The examples below are provided for illustrative purposes and are not intended to limit the scope of the present invention.

I. Detection of HCV RNA from Serum

RNA was extracted from serum using guanidinium salt, phenol and chloroform according to the

- instructions of the kit manufacturer (RNAzol B kit, Cinna/Biotecx). Extracted RNA was precipitated with isopropanol and washed with ethanol. A total of 25 µl serum was processed for RNA isolation, and the purified RNA was resuspended in 5 µl diethyl
- 20 pyrocarbonate treated water for subsequent cDNA synthesis.

II. <u>cDNA Synthesis and Polymerase Chain Reaction (PCR)</u> Amplification

Table 1 lists the sequence and position (with reference to HCV1) of all the PCR primers and probes used in these examples. Letter designations for

nucleotides are consistent with 37 C.F.R. §\$1.821-1.825. Thus, the letters A, C, G, T, and U are used in the ordinary sense of adenine, cytosine, quanine, thymine, and uracil. The letter M means A or C; R 5 means A or G; W means A or T/U; S means C or G; Y means C or T/U; K means G or T/U; V means A or C or G, not T/U; H means A or C or T/U, not G; D means A or G or T/U, not C; B means C or G or T/U, not A; N means (A or C or G or T/U) or (unknown or other). Table 1 is set forth below:

Table 1 Nucleotide Position Seq. No. Sequence (5'-3')

	_		
	67	CAAACGTAACACCAACCGRCGCCCACAGG	374-402
	•	ACAGAYCCGCAKAGRTCCCCCACG	1192-1169
15	68		509-538
	69	GCAACCTCGAGGTAGACGTCAGCCTATCCC	E00 E20
	70	GCAACCTCGTGGAAGGCGACAACCTATCCC	509-538
	• •	GTCACCAATGATTGCCCTAACTCGAGTATT	948-977
	71		948-973
	72	GTCACGAACGACTGCTCCAACTCAAG	
	•	TGGACATGATCGCTGGWGCYCACTGGGG	1375-1402
20	73		1375-1402
	74	TGGAYATGGTGGYGGGGGCYCACTGGGG	
	75	ATGATGAACTGGTCVCCYAC	1308-1327
	15	= :	1453-1428
	76	ACCTTVGCCCAGTTSCCCRCCATGGA	205-226
	77	AACCCACTCTATGYCCGGYCAT	203-220
	• •	GAATCGCTGGGGTGACCG	171-188
25	78		30-57
	79	CCATGAATCACTCCCCTGTGAGGAACTA	<u> </u>
	-	TTGCGGGGCACGCCCAA	244-227
	80	TIRCROGORIOGE	

25

For cDNA synthesis and PCR amplification, a protocol developed by Perkin-Elmer/Cetus (GeneAmp® RNA PCR kit) was used. Both random hexamer and primers with specific complementary sequences to HCV were employed to prime the reverse transcription (RT) reaction. All processes, except for adding and mixing reaction components, were performed in a thermal cycler (MJ Research, Inc.). The first strand cDNA synthesis reaction was inactivated at 99°C for 5 min, and then cooled at 50°C for 5 min before adding reaction 10 components for subsequent amplification. After an initial 5 cycles of 97°C for 1 min, 50°C for 2 min, and 72°C for 3 min, 30 cycles of 94°C for 1 min, 55°C for 2 min, and 72°C for 3 min followed, and then a final 7 min of elongation at 72°C. 15

Por the genotyping analysis, sequences 67 and 68 were used as primers in the PCR reaction. These primers amplify a segment corresponding to the core and envelope regions. After amplification, the reaction products were separated on an agarose gel and then transferred to a nylon membrane. The immobilized reaction products were allowed to hybridize with a 32p-labelled nucleic acid corresponding to either Genotype I (core or envelope 1) or Genotype II (core or envelope 1). Nucleic acid corresponding to Genotype 1 comprised sequences numbered 69 (core), 71 (envelope), and 73 (envelope). Nucleic acid corresponding to

15

20

Genotype II comprised sequences numbered 70 (core), 72 (envelope), and 74 (envelope).

The Genotype I probes only hybridized to the product amplified from isolates which had Genotype I sequence. Similarly, Genotype II probes only hybridized to the product amplified from isolates which had Genotype II sequence.

In another experiment, PCR products were generated using sequences 79 and 80. The products were analyzed as described above except Sequence No. 73 was used to detect Genotype I, Sequence No. 74 was used to detect Genotype II, Sequence No. 77 (5'UT) was used to detect Genotype III, and Sequence No. 78 (5'UT) was used to detect Genotype IV. Each sequence hybridized in a genotype specific manner.

III. Detection of HCV GI-GIV using a sandwich hybridization assay for HCV RNA

An amplified solution phase nucleic acid sandwich hybridization assay format is described in this example. The assay format employs several nucleic acid probes to effect capture and detection. A capture probe nucleic acid is capable of associating a complementary probe bound to a solid support and HCV nucleic acid to effect capture. A detection probe nucleic acid has a first segment (A) that binds to HCV nucleic acid and a second segment (B) that hybridizes to a second amplifier nucleic acid.

25

Label

The amplifier nucleic acid has a first segment (B*) that hybridizes to segment (B) of the probe nucleic acid and also comprises fifteen iterations of a segment (C). Segment C of the amplifier nucleic acid is capable of hybridizing to three labeled nucleic acids.

Nucleic acid sequences which correspond to nucleotide sequences of the envelope 1 gene of Group I HCV isolates are set forth in sequences numbered 81-99. Table 2 sets forth the area of the HCV genome to which the nucleic acid sequences correspond and a preferred use of the sequences.

Table 2 Complement of Probe Type Sequence No. Nucleotide Numbers 15 879-911 Label 81 912-944 Label 82 945-977 Capture 83 978-1010 Label 84 20 1011-1043 Label 85 1044-1076 Label 86 1077-1109 Label 87 1110-1142 Capture 88 1143-1175

SUBSTITUTE SHEET

89

- 40 -

Table 2 continued

	Probe Type	Sequence No.	Complement of Nucleotide Numbers
5		:	1176-1208
	Label	90	
	Label	91	1209-1241
	Label	92	1242=1274
	Capture	93	1275–1307
10	Label	94	1308-1340
	Label	95	1341–1373
	Label	96	1374-1406
	Label	97	1407-1439
	Capture	98	1440-1472
15	Label	99	1473-1505

Nucleic acid sequences which correspond to
nucleotide sequences of the envelope 1 gene of Group II
HCV isolates are set forth in sequences 100-118. Table
3 sets forth the area of the HCV genome to which the
nucleic acid corresponds and the preferred use of the
sequences.

- 41 -

Table 3

	Probe Type	Sequence No.	Complement of Nucleotide Numbers
5	Label	100	879 - 911
	Label	101	912-944
	Capture	102	945-977
	Label	103	978-1010
10	Label	104	1011-1043
	Label	105	1044-1076
	Label	106	1077-1109
	Capture	107	1110-1142
	Label	108	1143-1175
15	Label	109	1176-1208
	Label	110	1209-1241
	Label	111	1242=1274
	Capture	112	1275-1307
	Label	113	1308-1340
20	Label	114	1341-1373
	Label	115	1374-1406
	Label	116	1407-1439
	Capture	117	1440-1472
	Label	118	1473-1505
25			

Nucleic acid sequences which correspond to nucleotide sequences in the C gene and the 5'UT region

are set forth in sequences 119-145. Table 4 identifies the sequence with a preferred use.

Table 4

5		
	Probe Type	Sequence No.
	Capture	119
	Label	120
10	Label	121
,	Label	122
	Capture	123
	Label	124
	Label	125
15	Label	126
	Capture	127
	Label	128
	Label	129
	Label	130
20	Capture	131
	Label	132
	Label	133
	Label	134
	Label	135
25	Capture	136
	Label	137
	Label	138

Table 4 continued

	Probe Type	Sequence No.
5	Label	139
	Capture	140
	Label	141
	Label	142
	Label	143
10	Capture	144
	Label	145

The detection and capture probe HCV-specific segments, and their respective names as used in this assay were as follows.

Capture sequences are sequences numbered 119-122 and 141-144.

Detection sequences are sequences numbered 119-140.

Each detection sequence contained, in addition to
the sequences substantially complementary to the HCV
sequences, a 5' extension (B) which extension (B) is
complementary to a segment of the second amplifier
nucleic acid. The extension (B) sequence is identified
in the Sequence Listing as Sequence No. 146, and is
reproduced below.

AGGCATAGGACCCGTGTCTT

25

Each capture sequence contained, in addition to the sequences substantially complementary to HCV sequences, a sequence complementary to DNA bound to a solid phase. The sequence complementary to DNA bound to a solid support was carried downstream from the capture sequence. The sequence complementary to the DNA bound to the support is set forth as Sequence No. 147 and is reproduced below.

CITCTTTGGAGAAAGTGGTG

Microtiter plates were prepared as follows. White Microlite 1 Removawell strips (polystyrene microtiter plates, 96 wells/plate) were purchased from Dynatech Inc.

Each well was filled with 200 µl 1 N HCl and incubated at room temperature for 15-20 min. The plates were then washed 4 times with 1X PBS and the wells aspirated to remove liquid. The wells were then filled with 200 µl 1 N NaOH and incubated at room temperature for 15-20 min. The plates were again washed 4 times with 1X PBS and the wells aspirated to remove liquid.

Poly(phe-lys) was purchased from Sigma Chemicals, Inc. This polypeptide has a 1:1 molar ratio of phe:lys and an average m.w. of 47,900 gm/mole. It has an average length of 309 amino acids and contains 155 amines/mole. A 1 mg/ml solution of the polypeptide was mixed with 2M NaCl/lX PBS to a final concentration of

10

15

20

25

0.1 mg/ml (pH 6.0). A volume of 200 µl of this solution was added to each well. The plate was wrapped in plastic to prevent drying and incubated at 30°C overnight. The plate was then washed 4 times with 1X PBS and the wells aspirated to remove liquid.

The following procedure was used to couple the nucleic acid, a complementary sequence to Sequence No. 147, to the plates, hereinafter referred to as immobilized nucleic acid. Synthesis of immobilized nucleic acid having a sequence complementary to Sequence No. 133 was described in EPA 883096976. A quantity of 20 mg disuccinimidyl suberate was dissolved in 300 µl dimethyl formamide (DMF). A quantity of 26 OD₂₆₀ units of immobilized nucleic acid was added to 100 µl coupling buffer (50 mM sodium phosphate, pH 7.8). The coupling mixture was then added to the DSS-DMF solution and stirred with a magnetic stirrer for 30 min. An NAP-25 column was equilibrated with 10 mM sodium phosphate, pH 6.5. The coupling mixture DSS-DMF solution was added to 2 ml 10 mM sodium phosphate, pH 6.5, at 4°C. The mixture was vortexed to mix and loaded onto the equilibrated NAP-25 column. DSS-activated immobilized nucleic acid DNA was eluted from the column with 3.5 ml 10 mM sodium phosphate, pH 6.5. A quantity of 5.6 OD 260 units of eluted DSS-activated immobilized nucleic acid DNA was added to 1500 ml 50 mM sodium phosphate, pH 7.8. A volume of 50

10

25

µl of this solution was added to each well and the plates were incubated overnight. The plate was then washed 4 times with 1X PBS and the wells aspirated to remove liquid.

Final stripping of plates was accomplished as follows. A volume of 200 μ l of 0.2N NaOH containing 0.5% (w/v) SDS was added to each well. The plate was wrapped in plastic and incubated at 65°C for 60 min. The plate was then washed 4 times with 1X PBS and the wells aspirated to remove liquid. The stripped plate was stored with desiccant beads at 2-8°C.

Serum samples to be assayed were analyzed using PCR followed by sequence analysis to determine the genotype.

15 Sample preparation consisted of delivering 50 μl of the serum sample and 150 μl P-K Buffer (2 mg/ml proteinase K in 53 mM Tris-HCl, pH 8.0/0.6 M NaCl/0.06 M sodium citrate/8 mM EDTA, pH 8.0/1.3%SDS/16μg/ml sonicated salmon sperm DNA/7% formamide/50 fmoles capture probes/160 fmoles detection probes) to each well. Plates were agitated to mix the contents in the well, covered and incubated for 16 hr at 62°C.

After a further 10 minute period at room temperature, the contents of each well were aspirated to remove all fluid, and the wells washed 2X with washing buffer (0.1% SDS/0.015 M NaCl/ 0.0015 M sodium citrate). The amplifier nucleic acid was then added to

15

20

25

each well (50 µl of 0.7 fmole/µl solution in 0..48 M NaCl/0.048 M sodium citrate/0.1% SDS/0.5% "blocking reagent" (Boehringer Mannheim, catalog No. 1096 176)). After covering the plates and agitating to mix the contents in the wells, the plates were incubated for 30 min. at 52°C.

After a further 10 min period at room temperature, the wells were washed as described above.

Alkaline phosphatase label nucleic acid, disclosed in EP 883096976, was then added to each well (50 µl/well of 2.66 fmoles/µl). After incubation at 52°C for 15 min., and 10 min. at room temperature, the wells were washed twice as above and then 3X with 0.015 M NaCl/0.0015 M sodium citrate.

An enzyme-triggered dioxetane (Schaap et al., Tet. Lett. (1987) 28:1159-1162 and EPA Pub. No. 0254051), obtained from Lumigen, Inc., was employed. A quantity of 50 µl Lumiphos 530 (Lumigen) was added to each well. The wells were tapped lightly so that the reagent would fall to the bottom and gently swirled to distribute the reagent evenly over the bottom. The wells were covered and incubated at 37°C for 20-40 min.

Plates were then read on a Dynatech ML 1000 luminometer. Output was given as the full integral of the light produced during the reaction.

The assay positively detected each of the serum samples, regardless of genotype.

- 48 -

IV. Expression of the Polypeptide Encoded in Sequences Defined by Differing Genotypes

HCV polypeptides encoded by a sequence within sequences 1-66 are expressed as a fusion polypeptide with superoxide dismutase (SOD). A cDNA carrying such sequences is subcloned into the expression vector psoDcfl (Steimer et al. 1986)).

First, DNA isolated from pSODcfl is treated with BamHI and EcoRI, and the following linker was ligated into the linear DNA created by the restriction enzymes:

5 GAT CCT GGA ATT CTG ATA AGA

CCT TAA GAC TAT TIT AA 3
After cloning, the plasmid containing the insert is isolated.

Plasmid containing the insert is restricted with EcoRI. The HCV cDNA is ligated into this EcoRI linearized plasmid DNA. The DNA mixture is used to transform E. coli strain D1210 (Sadler et al. (1980)). Polypeptides are isolated on gels.

20

25

10

V. Antigenicity of Polypeptides

The antigenicity of polypeptides formed in Section IV is evaluated in the following manner. Polyethylene pins arranged on a block in an 8 12 array (Coselco Mimetopes, Victoria, Australia) are prepared by placing the pins in a bath (20% v/v piperidine in dimethylformamide (DMF)) for 30 minutes at room

10

15

20

25

temperature. The pins are removed, washed in DMF for 5 minutes, then washed in methanol four times (2 min/wash). The pins are allowed to air dry for at least 10 minutes, then washed a final time in DMF (5Min). 1-Hydroxybenzotriazole (HOBt, 367 mg) is dissolved in DMF (80 µL) for use in coupling Fmoc-protected polypeptides prepared in Section IV.

The protected amino acids are placed in micro-titer plate wells with HOBt, and the pin block placed over the plate, immersing the pins in the wells. The assembly is then sealed in a plastic bag and allowed to react at 25°C for 18 hours to couple the first amino acids to the pins. The block is then removed, and the pins washed with DMF (2 min.), MeOH (4 x, 2 min.), and again with DMF (2 min.) to clean and deprotect the bound amino acids. The procedure is repeated for each additional amino acid coupled, until all octamers are prepared.

The free N-termini are then acetylated to compensate for the free amide, as most of the epitopes are not found at the N-terminus and thus would not have the associated positive charge. Acetylation is accomplished by filling the wells of a microtiter plate with DMF/acetic anhydride/triethylamine (5:2:1 v/v/v) and allowing the pins to react in the wells for 90 minutes at 20°C. The pins are then washed with DMF (2

15

min.) and MeOH (4 x, 2 min.), and air dried for at least 10 minutes.

The side chain protecting groups are removed by treating the pins with trifluoroacetic acid/phenol/dithioethane (95:2.5:1.5, v/v/v) in polypropylene bags for 4 hours at room temperature. The pins are then washed in dichloromethane (2 x, 2 min.), 5% di-isopropylethylamine/dichloromethane (2 x, 5 min.), dichloromethane (5 min.), and air-dried for at least 10 minutes. The pins are then washed in water (2 min.), MeOH (18 hours), dried in vacuo, and stored in sealed plastic bags over silica gel. IV.B.15.b Assay of Peptides.

Octamer-bearing pins are treated by sonicating for 30 minutes in a disruption buffer (1% sodium dodecylsulfate, 0.1% 2-mercaptoethanol, 0.1 M NaH2PO4) at 60°C. The pins are then immersed several times in water (60°C), followed by boiling MeOH (2 min.), and allowed to air dry.

The pins are then precoated for 1 hour at 25°C in microtiter wells containing 200 μL blocking buffer (1% ovalbumin, 1% BSA, 0.1% Tween, and 0.05% NaN3 in PBS), with agitation. The pins are then immersed in microtiter wells containing 175 μL antisera obtained from human patients diagnosed as having HCV and allowed to incubate at 4°C overnight. The formation of a complex between polyclonal antibodies of the serum and

the polypeptide initiates that the peptides give rise to an immune response in vivo. Such peptides are candidates for the development of vaccines.

Thus, this invention has been described and illustrated. It will be apparent to those skilled in the art that many variations and modifications can be made without departing from the purview of the appended claims and without departing from the teaching and scope of the present invention.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- 5 (i) APPLICANT: Tai-An Cha
 - (ii) TITLE OF INVENTION: HCV GENOMIC SEQUENCES FOR DIAGNOSTICS AND THERAPEUTICS
- 10 (iii) NUMBER OF SEQUENCES: 147
 - (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
 - (B) STREET: 600 Atlantic Avenue
- 15 (C) CITY: Boston
 - (D) STATE: Massachusetts
 - (E) COUNTRY: USA
 - (F) ZIP: 02210
- 20 (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Diskette, 5.25 inch
 - (B) COMPUTER: IBM compatible
 - (C) OPERATING SYSTEM: MS-DOS Version 3.3
 - (D) SOFTWARE: WordPerfect 5.1

- 104 -

٠		TCTCTATCTT CCTCTTGGCT CTGCTGTCC		
	(2)	INFORMATION FOR SEQ ID NO: 60		
5		(i) SEQUENCE CHARACTERISTICS:		
		(A) LENGTH: 549 nucleotides		
		(B) TYPE: nucleic acid		
		(C) STRANDEDNESS: single		
		(D) TOPOLOGY: linear		
10		• •		
		(ii) MOLECULE TYPE: DNA		
		(vi) ORIGINAL SOURCE:		
		(C) INDIVIDUAL ISOLATE: nac5		
15				
		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 60		
		ATGAGCACAA ATCCTAAACC CCAAAGAAAA ACCAAACGTA		
		ACACCAACCG TCGCCCACAG GACGTCAAGT TCCCGGGCGG	80	
		TGGTCAGATC GTTGGTGGAG TTTACCTGTT GCCGCGCAGG	120	
20		GGCCCCAGGT TGGGTGTGCG CGCGACTAGG AAGACTTCCG	160	
		AGCGGTCGCA ACCTCGTGGA AGGCGACAAC CTATCCCCAA	200	
		GGCTCGCCGG CCCGAGGGCA GGTCCTGGGC TCAGCCCGGG	240	
		TACCCTTGGC CCCTCTATGG CAACGAGGGT ATGGGGTGGG		
		CAGGATGGCT CCTGTCACCC CGCGGCTCCC GGCCTAGTTG		
25		GGGCCCCACG GACCCCCGGC GTAGGTCGCG TAATTTGGGT		
_•		ARCOTOATOG ATACCOTOAC ATGCGGCTTC GCCGACCTCA		

- 105 -

	•	TGGGGTACAT TCCGCTCGTC GGCGCCCCCC TAGGGGGCGC	440
		TGCCAGGGCC CTGGCACATG GTGTCCGGGT TCTGGAGGAC	480
		GGCGTGAACT ATGCAACAGG GAATTTGCCT GGTTGCTCTT	520
		TCTCTATCTT CCTCTTGGCT CTGCTGTCC	549
5			
	(2)	INFORMATION FOR SEQ ID NO: 61	
		(i) SEQUENCE CHARACTERISTICS:	
		(A) LENGTH: 549 nucleotides	
		(B) TYPE: nucleic acid	
10		(C) STRANDEDNESS: single	
		(D) TOPOLOGY: linear	
		(ii) MOLECULE TYPE: DNA	
15		(vi) ORIGINAL SOURCE:	
		(C) INDIVIDUAL ISOLATE: arg2	
		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 61	
		ATGAGCACGA ATCCTAAACC TCAAAGAAAA ACCAAACGTA	40
20		ACACCAACCG CCGCCCACAG GACGTCAAGT TCCCGGGCGG	80
	-	TGGTCAGATC GTTGGTGGAG TTTACTTGTT GCCGCGCAGG	120
		GGCCCCAGGT TGGGTGTGCG CGCGACTAGG AAGACTTCCG	160
		AGCGGTCGCA ACCTCGTGGA AGGCGACAAC CTATCCCCAA	200
		GGCTCGCCAG CCCGAGGGTA GGGCCTGGGC TCAGCCCGGG	240
25		TACCCTTGGC CCCTCTATGG CAATGAGGGT ATGGGGTGGG	280
		CAGGGTGGCT CCTGTCCCCC CGCGGCTCCC GGCCTAGTTG	320

		(ii)	MOLECULE TYPE: DNA	
			SEQUENCE DESCRIPTION: SEQ ID NO: 145	
5		CARRA	GGAAG AKAGAGAAAG AGCAACCRGG MAR	33
	(2)	INFOR	MATION FOR SEQ ID NO: 146	
		(i)	SEQUENCE CHARACTERISTICS:	
10			(A) LENGTH: 20 nucleotides	
			(B) TYPE: nucleic acid	
		. <u>.</u>	(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	
15		(ii)	MOLECULE TYPE: DNA	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 146	
		AGGCA	TAGGA CCCGTGTCTT	20
20	(2)	INFOR	MATION FOR SEQ ID NO: 147	
		(i)	SEQUENCE CHARACTERISTICS:	
			(A) LENGTH: 20 nucleotides	
			(B) TYPE: nucleic acid	
25			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	
		(ii)	MOLECULE TYPE: DNA	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 147	
		CTTCT	ITGGA GAAAGTGGTG	20

- 150 -

CLAIMS

- 1. As a composition of matter, a non-naturally occurring nucleic acid having a non-HCV-1 nucleotide sequence of eight or more nucleotides corresponding to a nucleotide sequence within the hepatitis C virus genome.
- The composition of claim 1 wherein said nucleotide
 sequence corresponding to a non-HCV-1 nucleotide
 sequence within the hepatitis C virus genome is
 selected from the regions consisting of the NS5 region,
 envelope 1 region, 5'UT region, and the core region.
- 15 3. The composition of claim 1 wherein said nucleotide sequence corresponding to a non-HCV-1 nucleotide sequence within the hepatitis C virus genome corresponds to a sequence in the NS5 region.
- 20 4. The composition of claim 3 wherein said nucleotide sequence corresponding to a non-HCV-1 sequence within the hepatitis C virus genome is selected from a sequence within sequences numbered 2-22.

5. The composition of claim 1 wherein said nucleotide sequence corresponding to a non-HCV-1 nucleotide sequence within the hepatitis C virus genome corresponds to a sequence in the envelope 1 region.

5

6. The composition of claim 5 wherein said nucleotide sequence corresponding to a non-HCV-1 sequence within the hepatitis C virus genome corresponds to a sequence within sequence numbers 24-32.

10

7. The composition of claim 1 wherein at least one sequence corresponding to a non-HCV-1 nucleotide sequence within the hepatitis C virus genome corresponds to a sequence in the 5'UT region.

15

8. The composition of claim 7 wherein said nucleotide sequence corresponding to a non-HCV-1 sequence within the hepatitis C virus genome corresponds to a sequence within sequences numbered 34-51.

20

9. The composition of claim 1 wherein said nucleotide sequence corresponding to a non-HCV-1 nucleotide sequence within the hepatitis C virus genome corresponds to a sequence in the core region.

10. The composition of claim 9 wherein said nucleotide sequence corresponding to a non-HCV-1 sequence within the hepatitis C virus genome corresponds to a within sequences numbered 53-66.

5

11. The composition of claim 1 wherein said non-naturally occurring nucleic acid has a nucleotide sequence corresponding to one or more genotypes of hepatitis C virus.

10

- 12. The composition of claim 11 wherein said non-naturally occurring nucleic acid has a sequence corresponding to a sequence of a first genotype which first genotype is defined substantially by sequences numbered 1-6 in the NS5 region, 23-25 in the envelope 1 region, 33-38 in the 5'UT region, and 52-57 in the core region.
- 13. The composition of claim 11 wherein said
 20 non-naturally occurring nucleic acid has a sequence
 corresponding to a sequence of a second genotype which
 second genotype is defined substantially by sequences
 numbered 7-12 in the NS5 region, 26-28 in the envelope
 1 region, 39-45 in the 5'UT region, and 58-64 in the
 25 core region.

- 14. The composition of claim 11 wherein said non-naturally occurring nucleic acid has a sequence corresponding to a sequence of a third genotype which third genotype is defined substantially by sequences numbered 13-17 in the NS5 region, 32 in the envelope 1 region, 46-47 in the 5'UT region and 65-66 in the core region.
- 15. The composition of claim 11 wherein said
 10 non-naturally occurring nucleic acid has a sequence
 corresponding to a sequence of a fourth genotype which
 fourth genotype is defined substantially by sequences
 numbered 20-22 in the NS5 region, 29-31 in the envelope
 1 region and 48-49 in the 5'UT region.
- 16. The composition of claim 11 wherein said non-naturally occurring nucleic acid has a sequence corresponding to a sequence of a fifth genotype which fifth genotype is defined substantially by sequences numbered 18-19 in the NS5 region and 50-51 in the 5'UT region.
- 17. The composition of claim 1 wherein said non-naturally occurring nucleic acid is capable of
 25 priming a reaction for the synthesis of nucleic acid to form a nucleic acid having a nucleotide sequence corresponding to hepatitis C virus.

- 18. The composition of claim 1 wherein said non-naturally occurring nucleic acid has label means for detecting a hybridization product.
- 5 19. The composition of claim 1 wherein said non-naturally occurring nucleic acid has support means for separating a hybridization product from solution.
- 20. The composition of claim 1 wherein said
 10 non-naturally occurring nucleic acid prevents the
 transcription or translation of viral nucleic acid.
 - 21. A method of forming a hybridization product with a hepatitis C virus nucleic acid comprising the following steps:
- a. placing a non-naturally occurring nucleic acid having a nucleotide sequence of eight or more nucleotides corresponding to a non-HCV-1 sequence in the hepatitis C viral genome into conditions in which hybridization conditions can be imposed said non-naturally occurring nucleic acid capable of forming a hybridization product with said hepatitis C virus nucleic acid under hybridization conditions; and

20

- b. imposing hybridization conditions to form a hybridization product in the presence of hepatitis C virus nucleic acid.
- 5 22. The method of claim 21 wherein said nucleotide sequence corresponding to a non-HCV-1 sequence in the hepatitis C virus genome corresponds to a sequence within at least one of the regions consisting essentially of NS5 region, envelope 1 region, 5'UT region, and the core region.
 - 23. The method of claim 21 wherein said nucleotide sequence corresponds to a non-HCV-1 sequence corresponds to a sequence within the NS5 region.
 - 24. The method of claim 23 wherein said nucleotide sequence corresponds to a non-HCV-1 sequence corresponds to a sequence within sequences numbered 2-22.
 - 25. The method of claim 21 wherein said nucleotide sequence corresponds to a non-HCV-1 sequence corresponds to a sequence within the envelope 1 region.

25

- 26. The method of claim 25 wherein said nucleotide sequence corresponds to a non-HCV-1 sequence is selected from a sequence within sequences numbered 24-32.
- 27. The method of claim 21 wherein said nucleotide sequence corresponds to a non-HCV-1 sequence corresponding to a sequence within the 5'DT region.
- 10 28. The method of claim 27 wherein said nucleotide sequence corresponds to a non-HCV-1 sequence selected from a sequence within sequences numbered 34-51.
- 29. The method of claim 21 wherein said nucleotide 15 sequence corresponds to a non-HCV-1 sequence corresponding to a sequence within the core region.
- 30. The method of claim 29 wherein said nucleotide sequence corresponds to a non-HCV-1 sequence selected from a sequence within sequences numbered 53-66.
 - 31. The method of claim 21 wherein said nucleotide sequence corresponds to a non-HCV-1 nucleotide sequence corresponding to one or more genotypes of hepatitis C virus.

- 32. The method of claim 21 wherein said non-naturally occurring nucleic acid has a sequence corresponding to a sequence of a first genotype which first genotype is defined substantially by sequences numbered 1-6 in the NS5 region, 23-25 in the envelope 1 region, 33-38 in the 5'UT region, and 52-57 in the core region.
- 33. The method of claim 21 wherein said non-naturally occurring nucleic acid has a sequence corresponding to a sequence of a second genotype which second genotype is defined substantially by sequences numbered 7-12 in the NS5 region, 26-28 in the envelope 1 region, 39-45 in the 5'UT region, and 58-64 in the core region.
- 15 34. The method of claim 21 wherein said non-naturally occurring nucleic acid has a sequence corresponding to a sequence of a third genotype which third genotype is defined substantially by sequences numbered 13-17 in the NS5 region, 32 in the envelope 1 region, 46-47 in the 5'UT region and 65-66 in the core region.
- 35. The method of claim 21 wherein said non-naturally occurring nucleic acid has a sequence corresponding to a sequence of a fourth genotype which fourth genotype 25 is defined substantially by sequences numbered 20-22 in the NS5 region, 29-31 in the envelope 1 region and 48-49 in the 5'UT region.

- 36. The method of claim 21 wherein said non-naturally occurring nucleic acid has a sequence corresponding to a sequence of a fifth genotype which fifth genotype is defined substantially by sequences numbered 18-19 in the NS5 region and 50-51 in the 5'UT region.
- 37. The method of claim 21 wherein said hybridization product is capable of priming a reaction for the synthesis of nucleic acid.
- 38. The method of claim 21 wherein said non-naturally occurring nucleic acid has label means for detecting a hybridization product.
- 15 39. The method of claim 21 wherein said non-naturally occurring nucleic acid has support means for separating the hybridization product from solution.
- 40. The method of claim 21 wherein said non-naturally occurring nucleic acid prevents the transcription or translation of viral nucleic acid.
- 41. As a composition of matter, a non-naturally occurring polypeptide corresponding to a non-HCV-1

 25 nucleotide sequence of nine or more nucleotides which sequence of nine or more nucleotides corresponds to a sequence within hepatitis C virus genomic sequences.

- 42. The composition of claim 41 wherein said non-HCV-1 sequence is selected from one of the regions consisting of MS5 region, envelope 1 region, and the core region.
- 5 43. The composition of claim 41 wherein said non-HCV-1 nucleotide sequence corresponds to a sequence in the NS5 region.
- 44. The composition of claim 43 wherein said non-HCV-1
 sequence is selected from a sequence within sequences numbered 2-22.
- 45. The composition of claim 41 wherein said non-HCV-1 sequence corresponds to a sequence in the envelope 1 region.
 - 46. The composition of claim 45 wherein said non-HCV-1 sequence is selected from a sequence within sequences numbered 24-32.
 - 47. The composition of claim 41 wherein said non-HCV-1 sequence corresponds to a sequence in the core region.
- 48. The composition of claim 47 wherein said non-HCV-1 sequence is selected from a sequence within sequences numbered 52-66.

- 49. The composition of claim 41 wherein said non-HCV-1 nucleotide sequence has a nucleotide sequence corresponding to one or more genotypes of hepatitis C virus.
- 50. The composition of claim 41 wherein said non-HCV-1 nucleotide sequence has a sequence corresponding to a sequence of a first genotype which first genotype is defined substantially by sequences numbered 1-6 in the NS5 region, 23-25 in the envelope 1 region, and 52-57 in the core region.
- 51. The composition of claim 41 wherein said non-HCV-1 nucleotide sequence has a sequence corresponding to a sequence of a second genotype which second genotype is defined substantially by sequences numbered 7-12 in the NS5 region, 26-28 in the envelope 1 region, and 58-64 in the core region.
- 52. The composition of claim 41 wherein said non-HCV-1 nucleotide sequence has a sequence corresponding to a sequence of a third genotype which third genotype is defined substantially by sequences numbered 13-17 in the NS5 region, 32 in the envelope 1 region, and 65-66 in the core region.

- 53. The composition of claim 41 wherein said non-HCV-1 nucleotide sequence has a sequence corresponding to a sequence of a fourth genotype which fourth genotype is defined substantially by sequences numbered 20-22 in the MS5 region, 29-31 in the envelope 1 region and 48-49 in the 5'UT region.
- 54. The composition of claim 41 wherein said non-HCV-1 nucleotide sequence has a sequence corresponding to a sequence of a fifth genotype which fifth genotype is defined substantially by sequences numbered 18-19 in the NS5 region and 50-51 in the 5'UT region.
- 55. The composition of claim 41 wherein said
 15 polypeptide is capable of generating an immune reaction in a host.
 - 56. An antibody capable of selectively binding to the composition of claim 41.
 - 57. A method of detecting one or more genotypes of hepatitis C virus comprising the following steps:
- a) placing a non-naturally occurring nucleic acid having a nucleotide sequence of eight or more
 25 nucleotides corresponding to one or more genotypes of hepatitis C virus under conditions where hybridization conditions can be imposed,

- b) imposing hybridization conditions to form a hybridization product in the presence of hepatitis
 C virus nucleic acid; and
- c) monitoring the non-naturally occurring nucleic acid for the formation of a hybridization product, which hybridization product is indicative of the presence of the genotype of hepatitis C virus.
- 58. The method of claim 57 wherein said non-naturally occurring nucleic acid has a sequence corresponding to a sequence of a first genotype which first genotype is defined substantially by sequences numbered 1-6 in the NS5 region, 23-25 in the envelope 1 region, 33-38 in the 5'UT region, and 52-57 in the core region.
- 59. The method of claim 57 wherein said non-naturally occurring nucleic acid has a sequence corresponding to a sequence of a second genotype which second genotype is defined substantially by sequences numbered 7-12 in the NS5 region, 26-28 in the envelope 1 region, 39-45 in the 5'UT region, and 58-64 in the core region.

- 60. The method of claim 57 wherein said non-naturally occurring nucleic acid has a sequence corresponding to a sequence of a third genotype which third genotype is defined substantially by sequences numbered 13-17 in the NS5 region, 32 in the envelope 1 region, 46-47 in the 5'UT region and 65-66 in the core region.
- 61. The method of claim 57 wherein said non-naturally occurring nucleic acid has a sequence corresponding to

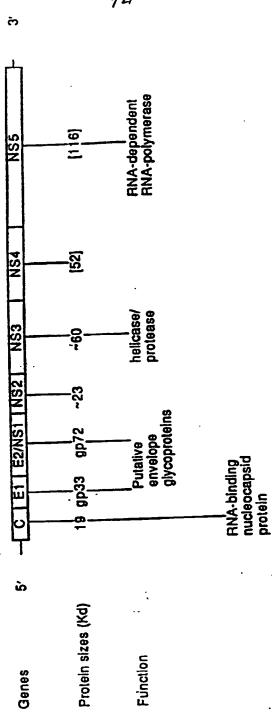
 10 a sequence of a fourth genotype which fourth genotype is defined substantially by sequences numbered 20-22 in the NS5 region, 29-31 in the envelope 1 region and 48-49 in the 5'UT region.
- 15 62. The method of claim 57 wherein said non-naturally occurring nucleic acid has a sequence corresponding to a sequence of a fifth genotype which fifth genotype is defined substantially by sequences numbered 18-19 in the WS5 region.
 - 63. The method of claim 57 wherein said non-naturally occurring nucleic acid has a sequence corresponding to a sequence numbered 67-145.

10

- 64. The method of claim 57 wherein said non-naturally occurring nucleic acid has a sequence corresponding to a sequence numbered 69, 71, 73 and 81-99 to identify Group I genotypes in the core and region of the HCV genome.
- 65. The method of claim 57 wherein said non-naturally occurring nucleic acid has a sequence corresponding to a sequence numbered 70, 72, 70 and 100-118 to identify Group II genotypes in the core and envelope regions of the HCV genome.
- 66. The method of claim 57 wherein said non-naturally occurring nucleic acid has a sequence corresponding to
 15 a sequence numbered 77 to identify Group III genotypes in the 5' UT region of the HCV genome.
- 67. The method of claim 57 wherein said non-naturally occurring nucleic acid has a sequence numbered 79 to identify Group IV genotypes in the 5' UT region of the HCV genome.

Fig.





SUBSTITUTE SHEET

2/21

Ø
~
•
ن
F
-

SEQUENCE 1 CTCCACAGTC ACTOAGAGG ACATCCGTAC GAAGAGGCA ATCTACCAT 2 1 CTCCACAGTC ACTOAGAGG ACATCCGTAC GAAGAGGCA ATCTACCAT 3 1 CTCCACAGTC ACTOAGAGG ACATCCGTAC GAAGAGGCA ATCTACCAT 4 1 CTCCACAGTC ACTOAGAGG ACATCCGTAC GAAGAGGCA ATCTACCAT 5 1 CTCCACAGTC ACTOAGAGG ACATCCGTAC GAAGAGGCA ATCTACCAT 6 1 CTCCACAGTC ACTOAGAGG ACATCCGTAC GAAGAGGCA ATCTACCAT 7 GII 1 CTCCACAGTC ACTGAAAGG ATATCCGTAC GAAGAGGCA ATTACCAT 8 1 CTCCACAGTC ACTGAAAAG ATATCCGTAC GAAGAGGCA ATTACCAT 10 CTCCACAGTC ACTGAAAAG ACATCCGTAC GAAGAGGCA ATTACCAAT 11 CTCCACAGTC ACTGAAAATG ACATCCGTAC TAAGAGGCA ATTACCAAT 12 CTCCACAGTC ACTGAAAATG ACATCCGTAC TAAGAGGCA ATTACCAAT 13 GIII 1 CTCCACAGTC ACTGAAATG ACATCCGTAT TAAGAAGTCA ATTACCAAT 14 CTCCACAGTC ACTGAAATG ACATCCGTAT TAAGAAGTCA ATTACCAAT 15 CTCCACAGTC ACTGAAATG ACATCCGTAT TAAGAAGTCA ATTACCAAT 16 CTCCACAGTC ACTGAAATG ACATCCGTAT TAAGAAGTCA ATTACCAAT 17 CTCCACAGTC ACTGAAAGG ACATCCGTAT TAAGAAGTCA ATTACCAAT 18 GII 1 CTCCACAGTC ACTGAAAGG ACATCCGTAT TAAGAAGTCA ATTACCAAT 19 CTCAACAGTC ACTGAAAGG ACATCCAGTC TAAGAAGTCC ATTACCAAT 10 CTCAACAGTC ACTGAAAGG ACATCACACA CATAACCAAT 11 CTCCACAGTC ACTGAAAGG ACATCACACA CATAACCAAT 12 CTCAACAGTC ACTGAAAGG ACATCACACA CATAACCAAT 13 GIII 1 CTCTACAGTC ACTGAAAGA CATAAAAAC CATAAAATCA ATTACCAAT 14 CTCAACAGTC ACTGAAAATGA CATAAAAAC TAAAAATCA ATTACCAAT 15 CTCAACAGTC ACAGAACAG ACATAAAAAC TAAAAATCA ATTACCAAT 16 CTCAACAGTC ACAGAACAG ACATAAAAAC TAAAAATCA ATTACCAAT 17 CTCAACAGTC ACAGAACAG ACATAAAAAC TAAAAATCA ATTACCAAT 18 GV I CTCCACAGTC ACAGAACAG ACATAAAAAC TAAAAACCAAT 19 CTCCACACTTA ACCGAACAAG ACATAAAAAC TAAAAACCAAT 10 CTCCACACTTA ACCGAACAAG ACATAAAAAC TAAAAACCAAT 11 CTCCACACTTA ACCGAACAAG ACATAAAAAC TAAAAACCAAT 11 CTCCACACTTA ACCGAACAAG ACATAAAAACA TAAAAACAAT ATTACCAATAAAAACAAT TAAACAAATAACAAT TAAACAAATAAAAAAAA	
R GENOTYPE SESCUENCES GII GII GII GIV GIV	
2	ACATCCGIAC GGAGGAGGCA ATCTACCAAT GTTGTGACCT CGACCCCCAA
5 6 6 6 7 7 7 6 9 9 9 9 10 11 11 11 11 11 11 11 11 11	ACATCCGIAC GGAGGAGGCA ATITACCAAT GINGIOACCT GGACCCCCAA
5 5 5 6 7 7 7 6 9 9 9 10 11 11 11 11 11 11 11 11 11	ACAICCGIAC GGAGGAGGCA AICIACCAAI GIIGIGAICI GGACCCCCAA
55 7 7 8 9 9 10 11 11 11 11 11 11 11 11 11	ACATCCGIAC GGAGGAGGCA ATTTACCAAT GTTGTGACCT GGACCCCCAA
6	ATATCCGTAC GGAGGAGGCA ATCTACCAGT GTTGTGACCT GGACCCCCAA
######################################	GIIGIGACCI G
8	eronoscopo
9 10 11 12 13 14 15 16 17 18 19 19 19 19 19 19	TOUCHUSTON AMERICAN AND AND AND AND AND AND AND AND AND A
9 10 11 12 13 14 15 16 17 18 19 19 19 19 19	TGAGGAGICA ATTIACCAAA GIIGIGACII
10 11 12 13 GIII 14 15 16 17 18 GV 19 19	TGAGGAGTCA ATTTATCAAT GTTGTGCCTT
11 12 13 GIII 14 15 16 17 18 GV 19 19	CGAGGAGICG AITTACCAAI GIIGIGACII
12 13 GIII 14 15 16 17 18 GV 19 19	
13 GIII 14 15 16 17 18 GV 19 19 19 19 19	ACAICCGIGI IGAGGAGICA AICIACCAAI GINGIGACIII GGCCCCCGAA
	ACATCAGAAC TGAGGAGTCC ATATACCGAG CCTGCTCCCT
	ACATCACATC CTAGGAGTCC ATCTACCAGT CCTGTTCACT GCCCGAGGAG
19 19	
	ACATAAGAAC AGAAGAATCC ATATATCAGG GTTGTTCCCT GCCTCAGGAG
	计算机 化二甲基苯甲甲甲甲甲甲甲甲甲甲甲甲甲甲甲甲甲甲甲甲甲甲甲甲甲甲甲甲甲甲甲甲甲甲甲
1 0 1 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1	ATTIACCAAT CATIGIACTI
5 C C	ACATAATGAC TGAAGAGTCC ATTTACCAAT CATTGTACTT GCAGCCTGAG
GIV 1 CTCTACTGTC	19 的复数计算 医多种性 医多种性 医多种性 医多种性 医多种性 医多种性 医多种性 医多种性
	GGAAGAGGAG ATATACCAGT GCTGTAACCT
21 1 CTCGACTGTC ACTGAACAGG ACATCAGGGT	GGAAGAGGAG ATATACCAAT GCTGTAACCT
22 1 CICAACIGIC ACIGAACAGG ACAICAGGGI	ACATCAGGGT GGAAGAGGAG ATATACCAAT GCTGTAACCT TGAACCGGAG

3/21

2L	
ш	
L L	

$\overline{}$
10
Ξ.
-
~
$\overline{}$
•
_
좆
ð
Ħ
6
22
œ
S
6
z

)
1			NES REGION - (2/5)
HOOD HAND CHARLES OF THE BEST	6 6 8 8 8	0 M M D	计小计划 经自己货售 医皮色白色 经帐篷 医拉拉氏征 医克里氏 医共享共享性 医马拉耳氏征 计自动设备 医复数医疗 医多种 医多种性 医多种性 医多种性 医多种性 医多种性 医多种性 医多种性
SECOENCE			
NUMBER	GENOTYPE		
		71	GCCCGCGTGG CCATCAAGTC CCTCACCGAG AGGCTTTATG TTGGGGGCCC TCTTACCAAT TCAAGGGGG
~	ij	11	CCATCAAGIC CCICACIGAG AGGCITTAIG
т	ម	7.1	OCCCOCCIO CCATCAAGIC CCICACIGAG AGCCITIACG TIGGGGGCCC ICTIACCAAI ICAAGGGGG
~	19	12	GCCCGCGTGG CCATCAAGTC CCTCACTGAG AGGCTTTATG TTGGGGGCCC CCTTACCAAT TCAAGGGGGG
ĸ	19	11	GCCCGCGTGG CCATCAAGTC CCTCACCGAG AGGCTTTATG TCGGGGGCCC TCTTACCAAT TCAAGGGGG
vo	10	71	GCCCGIGIGG CCAICAAGIC CCICACIGAG AGGCIIIAIG IIGGGGGCCC ICIIACCAAI ICAAGGGGGG
		84	ossesennamennandenmennandennandenmannamennandenmennandenmendenmennandenmendenmennandenmennandenmen. 71 GCC+GC+GC+GC CC+T+AGGTC GCTC+CAGAG CGGCTCTATG TCGGGGGTCC TATGACTAAC TCCAAAGGGC
- a	;	: [CCATAAGGTC GCTCACAGAG CGGCTTTACA TCGGGGGCCC CCTGACTAAT
.		; ;	CCATABAGGGC GCICACAG CGGCTTTAGA TCGGGGGGCC CCTGACGAG
		: 5	CONTINUE CONTINUE CONTINUE ALCOHOLING CONTINUE ACCORDING CONTINUE ACCORDINA ACCORDING CONTINUE ACCORDING CONTINUE ACCORDING CONTINUE ACCORDINA ACCORDING CONTINUE ACCORDINA A
2 -		: =	THE PROPERTY CONTRACTOR AND CONTRACTOR CONTRACTOR CONTRACTOR AND CONTRACTOR C
1 6		: 5	GENERAL COCTITION TEGEGGETE CETABLE
3		- 1	でもころでである。そのものこのでは、「「「「「「」」」」と、「「」」」」と、「「」」」」」」」、「「」」」」」、「「」」」」、「「」」」」、「「」」」、「「」」」、「「」」、「「」」、「「」」、「「」」、「「」」、「「」
		71	1
7		11	
15		7.1	GCCCGAACTG CIATACACTC ACTGACTGAG AGACTGTACG TAGGGGGGCC CATGACAAAC AGCAAGGGGC
91		11	GCTCGAACTG CCATACACTC ACTGACTGAG AGGCTGTACG TAGGGGGGCC CATGACAAAC AGCAAAGGGC
11		11	GCTAGAACTG CTATCCACTC GCTCACTGAG AGACTCTACG TAGGAGGGCC CATGACAAAC AGCAAĞGGAC
8 0 0 0 0 0 0	8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8	A	REGERTATION CONTINUES CONT
æ 6 → ,	3	7 ;	でいらいははいなく じんみじゅうしゅうしょう でんじんきん でんじじょうしゅう アンドラ アンドラング アンドラン アンドランジン アンドランジング アンドランジングランド アンドランジングランド アンドランジング アンドランジング
7 T			できましょう こくしょく さんしんしん しょうしん しょうしんじん くうしんしょう しょうしん しょうしゅう はんしゅう しょうしん しょうしゅう しゅうしゅう しゅう
20	1	,	
17	, ,	17	IGAICICCIC CCICACGGAG CGGCITIACI
22		11	OCCAGGAAAG TGATCTTCTT CCTCACGGAA CGGCTTTACT GCGGGGCCCC TATGTTCAAC AGCAAGGGG

Fig. 2

D
`
m
$\boldsymbol{\smile}$
중
7
≅
-
\sim
200
24
_
83

! !		51 15 17	19 13 13 15 13	11 11 11 11 11 11 11 11 11 11 11 11 11
D NUMBER GENOTYPE	CCCTCACTTG CCCTCACTTG CCCTCACTTG CCCTCACTTG CCCTCACTTG CCCTCACTTG	AGAACTGCGG CTATCGCCGG TGCCGCCCA GCGGCTGCT GACGACTAGC TGCGGTAATA CCCTCACATG AGAACTGCGG CTATCGCCGG TGCCGCCCA GCGGCTGCT GACGACTAGC TGCGGTAATA CCCTCACATG AGAACTGCGG TTATCGCCGG TGCCGCGCCA GCGGCTACT GACGACCAGC TGCGGTAATA CCCTTACATG AGAACTGCGG TTATCGCCGG TGCCGCGCA GCGGCGTGCT GACGACTAGC TGCGGTAATA CCCTCACATG AGAACTGCGG CTATCGCCGG TGCCGCGCA GCGGCGTGCT GACGACTAGC TGCGGTAATA CCCTCACATG AGAACTGCGG CTATCGCCGG TGCCGCGCAA GCGGCGTGCT GACGACTAGC TGCGGTAATA CCCTCACATG	AGACCTGCGG GTACAGGGGT TGCCGCGCG GCGGGTGCT CACCACTAGC ATGGGGAACA CCATCACATG AATCCTGCGG GTACAGGCGT TGCCGCGCGA GCGGGTGCT CACCACCAGC ATGGGCAACA CCATCACGTG AATCCTGCGG GTACAGGCGT TGCCGCGCGA GCGGAGTGCT CACCACCAGC ATGGGCAACA CGCTCACGTG AATCCTGCGG GTACAGGCGT TGCCGCGCGA GCGGAGTGCT CACCACCAGC ATGGGTAACA CACTCACGTG AATCCTGCGG TTACAGGCGT TGCCGCGCCA GCGGGGTCTT CACCACCAGC ATGGGGAATA CCATGACATG AACAATGTGG TTATCGTAGA TGCCGCGCCA GCGGCGTCTT CACCACTAGT ATGGGCAACA CCATGACGTG AACAATGTGG TTATCGTAGA TGCCGCGCCA GCGGCGTCTT CACCACCAGT ATGGGCAACA CCATGACGTG	141 CCCAGTGTGG TTATCGCCGT TGCCGTGCTA GTGGAGTCCT GCCTACCAGC TTCGGCAACA CAATCACTIG 141 CCCAGTGTGG TTATCGCCGT TGCCGTGCTA GTGGAGTCCT GCCTACCAGC TTCGGCAACA CAATCACTIG 141 CCCAGTGTGG TTATCGCCGT TGCCGTGCCA GTGGAGTTCT GCCTACCAGC TTCGGCAACA CAATCACTTG 141 CCCAGTGTGG TTATCGCCGT TGCCGTGCCA GTGGAGTTCT GCCTACCAGC TTCGGCAACA CAATCACTTG
8 8 8 8 8 8 8 8 8 8	CTATCGCAGG TGCCGCGCA GCGGCGTACT GACAACTAGC TGTGGTAACA CCCTCACTTG CTACCGCAGG TGCCGGCGA GCGGCGTACT GACAACTAGC TGTGGTAACA CCCTCACTTG CTACCGCAGG TGCCGGGCGA GCGGCGTACT GACAACTAGC TGTGGTAATA CCCTCACTTG CTATCGCAGG TGCCGCGGA GCGGCGTACT GACAACTAGC TGTGGTAACA CCCTCACTTG CTATCGCAGG TGCCGCGCAA GCGGCGTACT GACAACTAGC TGTGGTAACA CCCTCACTTG	TGCCGCGCAA GCGGCGTGCT GACGACTAGC TGCGGTAATA CCCTCACATG TGCCGCGCCA GCGGCTACT GACGACTAGC TGCGGTAATA CCCTCACATG TGCCGCGCCA GCGGCGTACT GACGACCAGC TGCGGTAATA CCCTTACATG TGCCGCGCGA GCGGCGTGCT GACGACTAGC TGCGGTAATA CCCTCACATG TGCCGCGCAA GCGGCGTGCT GACGACTAGC TGCGGTAATA CCCTCACATG	AGACCTGCGG GIACAGGCGT TGCCGCGCCA GCGGGGTGCT CACCACTAGC ATGGGGAACA CCATCACTAGGAACA CCATCACTAGAACACACACACACACACACACACACACACA	CCCAGTGIGG TIAICGCCGT TGCCGTGCTA GTGGAGTCCT GCCTACCAGC TTCGGCAACA CAATCACTGCCCAGTGGGGGGGGGG
	CTATCGCAGG TGCCGCGCA GCGGCGTACT GACAACTAGC TGTGGTAACA CTACCGCAGG TGCCGCGCGA GCGGCGTACT GACAACTAGC TGTGGTAACA CTACCGCAGG TGCCGGGCGA GCGGCGTACT GACAACTAGC TGTGGTAATA CTATCGCAGG TGCCGCGCGA GCGGCGTACT GACAACTAGC TGTGGTAACA CTATCGCAGG TGCCGCGCAA GCGGCGTACT GACAACTAGC TGTGGTAACA CTACCGCAGG TGCCGCGCAA GCGGCGTACT GACGACTAGC TGTGGTAATA	GACGACTAGC GACGACTAGC GACGACCAGC GACGACTAGC GACGACTAGC GACGACTAGC	CACCACTAGE CACCACCAGE CACCACCAGE CACCACCAGE CACCACCAGE CACCACCAGE	TTATCGCCGT TGCCGTGCTA GTGGAGTCT GCCTACCAGC TTATCGCCGT TGCCGTGCTA GTGGAGTTCT GCCTACCAGC TTATCGCCGT TGCCGTGCCA GTGGAGTTCT GCCTACCAGC
	TOCCGCGCGA GCGGCGTACT GACAACTAGC TGCCGGGCGA GCGGCGTACT GACAACTAGC TGCCGGGCGA GCGGCGTACT GACAACTAGC TGCCGCGCGA GCGGCGTACT GACAACTAGC TGCCGCGCAA GCGGCGTACT GACAACTAGC	CTATCOCCGO TOCCGCGCA GCGGCGTGCT GACGACTAGC CTATCGCCGO TOCCGCGCCA GCGGCGTACT GACGACTAGC TTATCGCCGG TGCCGCGCGA GCGGCGTGCT GACGACTAGC CTATCGCCGG TGCCGCGCGA GCGGCGTGCT GACGACTAGC CTATCGCCGG TGCCGCGCAA GCGGCGTGCT GACGACTAGC CTATCGCCGG TGCCGCGCAA GCGGCGTGCT GACGACTAGC	TGCCGCGCCA GCGGGGTGCT TGCCGCGCGA GCGGAGTGCT TGCCGCGCGA GCGGAGTGCT TGCCGCGCGA GCGGAGTGCT TGCCGCGCGA GCGGGGTCTT TGCCGCGCCA GCGGGCGTCTT TGCCGCGCCCA GCGGCGTCTT TGCCGCGCCCA GCGGCGTCTT TGCCGCGCCCA GCGGCGTCTT TGCCGCGCCCA GCGGCGTCTT TGCCGCGCCCA	GTGGAGTTCT GTGGAGTTCT GTGGAGTTCT GTGGAGTTCT
	TGCGGGGGA TGCGGGGGA TGCGGGGGA TGCGGGGGA TGCGGGGAA	TGCCGCGCA TGCCGCGCGA TGCCGCGCGA TGCCGCGCAA	TGCCGCGCGA TGCCGCGCGA TGCCGCGCGA TGCCGCGCGA TGCCGCGCCA TGCCGCGCCA	TOCCOTOCTA TOCCOTOCTA TOCCOTOCCA
	CTATCGCAGG CTACCGCAGG CTACCGCAGG CTATCGCAGG CTATCGCAGG	AGAACTGCGG CTATCGCCGG AGAACTGCGG TTATCGCCGG AGAACTGCGG TTATCGCCGG AGAACTGCGG TTATCGCCGG AGAACTGCGG CTATCGCCGG	GTACAGGCGT GTACAGGCGT GTACAGGCGT TTACAGGCGT TTACAGGCGT	TTATCGCCGT TTATCGCCGT TTATCGCCGT
	AGAACTGCGG AGAACTGCGG AGAACTGCGG AAAACTGCGG AAAACTGCGG	AGAACTGCGG AGAACTGCGG AGAACTGCGG AGAACTGCGG AGAACTGCGG	AGACCTGCGG AATCCTGCGG AATCCTGCGG AATCCTGCGG AATCCTGCGG AATCCTGCGG	141 CCCAGTGTGG 141 CCCAGTGTGG 141 CCCAGTGTGG 141 CCCAGTGTGG
; 6 5	14111411			141 141 141 141
GENOTYPE	10	III	13 GISI 14 15 16 17 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	20 GIV GIV 21 22 23 23 23 23 23 23 23 23 23 23 23 23
SEQUENCE ID NUMBER GENOTYPE	ସମାପ୍ୟସରେ	011 011 12 12 12 12 12 12 12 12 12 12 12 12 1	13 GISI 14 15 16 17 17 18 GV	

Fig. 2d

NSS REGION - (4/5)

SEQUENCE ID NUMBER GENOTYPE	Genotype			:					
0 6 7 8 9			i		CCTGTCGAGC CCTGTCGAGC CCTGTCGAGC CCTGTCGAGC CCTGTCGAGC	CCTGTCGAGC CGCAGGGCTC CCTGTCGAGG CGCAGGGCTC CCTGTCGAGG CGCAGGGCTC CCTGTCGAGC CGCAGGGCTC CCTGTCGAGC CGCAGGGCTC CCTGTCGAGC CGCAGGGCTC CCTGTCGAGC CGCAGGGCTC	CAGGACTGCA CAGGACTGCA CAGGACTGCA CAGGACTGCA CGGGACTGCA	CCATGCTCGT CCATGCTCGT CCATGCTCGT CCATGCTCGT CCATGCTCGT CCATGCTCGT	n
		n (CTACCTGAAG TTACTTGAAG TTACTTGAAG TTACTTGAAG TTACTTGAAG	GCCACAGGG GCCACTGCGG GCCTCTGCAG GCCTCTGCAG GCCTCTGCGG	CCTGTCGAGC CCTGTCGAGC CCTGTCGAGC CCTGTCGAGC CCTGTCGAGC	CCTGTCGAGC TGCCAAGCTC CCTGTCGAGC CGCGAAGCTC CCTGTCGAGC CGCGAAGCTC CCTGTCGAGC TGCAAGCTC CCTGTCGAGC TGCGAAGCTC	CAGGACTGCA CAGGACTGCA CAGGACTGCA CAGGACTGCA CAGGACTGCA CAGGACTGCA	CGATGCTCGT CGATGCTCGT CGATGCTCGT CGATGCTCGT CGATGCTCGT CAATGCTCGT	GAACGGAGAC GTGCGGAGAC GTGTGGGGAC GAACGGGGAC GTGCGGTGAC
ស្រុក្ស ព្រំក្រុ	13 GIII 14 15 16 17 GU	1 6	CIACOTAAAA GCCCTAGCGG CTTGCAGGC TGCAGGGAIA GTTGCACCCT CAATGCTGGT ATGCGGCGAC LI CTACGTAAA GCCAGGGCG CGTGTAACGC CGCGGGGAIT GTTGCTCCCA CCATGCTGGT GTGCGGCGAC LI CTACGTGAAA GCCAGAGCGG CGTGTAACGC CGCGGGCAIT GTTGCTCCCA CCATGTTGGT GTGCGGCGAC LI CTACGTGAAA GCTAAAGCGG CATGTAAACG CGCGGGCAIT GTTGCCCCCA CCATGTTGGT GTGCGGCGAC LI CTACATCAAA GCCTTGCAG CGTGCAAAGC TGCAGGGATC GTGGACCCTA TCATGCTGGT GTGTGGAGAC LI CTACATTAAG GCTTTAAGCT CCTGTAAAGC CGCAAAGCTC CAGGACTGCA CGCTCCTGGT GTGTGGTGAT	GCCCTAGCGG GCCAGGGCGG GCCAGAGCGG GCTAAAGCGG GCCTTGCAG	GCCCTAGCGG CTTGCAAGGC GCCAGGGCGG CGTGTAACGC GCCAGAGCGG CGTGTAACGC GCTAAAGCGG CATGTAACGC GCCTTGCAG CGTGCAAAGC	CTTGCAAGGC TGCAGGGATA CGTGTAACGC CGCGGGCATT CATGTAACGC CGCGGGCATT CATGTAACGC CGCGGGCATT CGTGCAAAGC TGCAAAGCTC	CTATOTAAAA GCCCTAGCGG CTTGCAAGGC TGCAGGGATA GTTGCACCCT CAATGCTGGT CTACGTAAAA GCCAGAGCGG CGTGTAAGGC CGCGGGGATT GTTGCTCCCA CCATGCTGGT CTACGTGAAA GCCAGAGCGG CGTGTAACGC CGCGGGCATT GTTGCTCCCA CCATGTTGGT CTACGTGAAA GCTAAAAGCGC CGCGGGCATT GTTGCCCCCA CCATGTTGGT CTACGTCAAAA GCTCAAAAGC TGCAGGGATC GTGGACCCTA TCATGCTGGT CTACATCAAAAG GCTTTAAACCT CCTGTAAAAGCTC CAGGACTGCA CGCTCCTGGT	CATGCTGGT CCATGCTGGT CCATGTTGGT CCATGTTGGT TCATGCTGGT	ATGCGCGAC GTGCGGCGAC GTGCGGCGAC GTGTGGAGAC
19 20 22 22	19 20 GIV 21 22	211 211 211 211 211		GCTTCAGCCG GCTTCAGCCG GCTAGAGCGG GCTAGAGCGG	GCTTCAGCCG CCTGTAGAGC TGCAAAGCTC GCTAGAGCGG CTTCGAAGGC CGCAGGCCTC GCTAGAGCGG CTGCGAAGGC CGCAGGCCTC GCTAGAGCGG CTGCGAAGGC CGCAGGGCTC	TGCAAAGCTC EEGGGGGCTC CGCAGGGCTC CGCAGGGCTC	CTACATCAAG GCTTCAGCCG CCTGTAGAGC TGCAAAGCTC CAGGACTGCA CGCTCCTGGT GTGTGGTGTG	CAGGACTGCA CGCTCCTGGT BELLEGIER BELLEGIER CGGAACCCGG ACTTTCTTGT CGGAACCCGG ACTTTCTTGT	GTGTGGTGTG CTGCGGAGAT CTGCGGAGAT CTGCGGAGAT

Fig. 26

NSS REGION - (5/5)

D NUMBER GENOTYPE	GTCG TTATCTGTGA (GTCG TTATCTGTGA	TIATCTGTGA AAGTGCGGG GTCCAGGAGG ACGCGGCGAG CCTGAGAGCC TTATCTGTGA AAGTGCGGG GTCCAGGAGG ACGCGGCGAG CCTGAGAGCC TTATCTGTGA AAGTGCGGG GTCCAGGAGG ACGCGGCGAG CCTGAGAGCC TTATCTGTGA AAGTCGGGG GTCCAGGAGG ACGCGGCGAA CTTGAGAGCC TTATCTGTGA AAGTCAGGGA GTCCAGGAGG ATGCAGGAGC CCTGAGAGCC TTATCTGTGA AAGTGCGGGG GTCCAGGAGG ATGCGGGCGAA CCTGAGAGCC TTATCTGTGA AAGTGCGGGG ACCCAGAGG ACGCGGCGAG CCTACGAGAGC TTATCTGTGA AAGTGCGGGGA ACCCAGAGG ACGCGGCGAA CCTACGAGAGC TTATCTGTGA AAGTGCGGGGA ACCCAGAGG ACGCGGCGAA CCTACGAGAGC TTATCTGTGA AAGTGCGGGGA ACCCAGAGG ACGCGGCGAA CCTACGAGTC TTATCTGTGA AAGTGCGGGGA ACCCAGAGG ACGCGGCGAA CCTACGAGTC TTATCTGTGA AAGTGCGGGGA ACCCAAGAGG ACGCGGCGAA CCTACGAGTC TTATCTGTGA AAGTGCGCGGAA ACCCAAGAGG ACGCGGCGAA CCTACGAGTC	GTCCAGGAGG GTCCAGGAGG GTCCAGGAGG GTCCAGGAGG GTCCAGGAGG ACCCAGGAGG ACCCAGGAGG ACCCAGGAGG ACCCAGGAGG ACCCAGGAGG	ACCCGCGAG ACCCGCGAG ACCCGCGAG ACCCGCGAA ACCCGCGAA ACCCGCCAA ACCCGCCAA ACCCGCCAA ACCCGCCAA ACCCGCCAA ACCCGCCAA	CCTGAGAGCC CCTGAGAGCC CCTGAGAGCC CCTGAGAGCC CCTGAGAGCC CCTGAGAGCC CCTGCGAGCC CCTACGAGCC CCTACGAGCC CCTACGAGCC CCTACGAGCC CCTACGAGCC CCTACGAGCC
1 GI 281 GACTTA 2 2 281 GACTTA 3 281 GACTTA 4 281 GACTTA 5 281 GACTTA 5 281 GACTTA 6 281 GACTTA 7 GII 281 GACTTA 8 281 GACTTA 8 281 GACTTA 10 281 GACTT 10 281 GACTT 11 281 GACTT 11 281 GACTT 12 281 GACTT 13 GII 281 GACTT 13 GII 281 GACTT 11 281 GACTT 12 281 GACTT 13 GII 281 GACTT 14 GACTT 15 GACTT 16 GACTT 17 GII 281 GACTT 18 GA	GTCG TIATCTGTGA GTCG TIATCTGTGA GTCG TIATCTGTGA GTCG TIATCTGTGA GTCG TIATCTGTGA GTCG TIATCTGTGA GTCG TIATCTGTGA GTCG TIATCTGTGA GTCG TIATCTGTGA GTCG TIATCTGTGA	AAGCGCGGG AAGCGCGGGA AAGCGCGGGG AAGCGCGGGG AAGCGCGGGG AAGCGCGGGA AAGCGCGGGGA AAGCGCGGGA AAGCGGA AAGCGCGGGA AAGCGGA AAGCGCGGA AAGCGCGGA AAGCGCGGA AAGCGCGGA AAGCGGA AAGCGCGGGA AAGCGGA AAGCGGA AAGCGGA AAGCGGA AAGCGGA AAGCGGA AAGCGGA AAGCGGA AAGCGA AAGCA AAGCGA AAGCA AAGCGA AAGCA AA	TECCAGGAGG ATCCAGGAGG ATCCAGGAGG ATCCAGGAGG ATCCAGGAGG ACCCAGGAGG ACCCAAGAGG ACCCAAGAGG ACCCAAGAGG	ACCCGCCAAAACCCGCGAAACCCGCGCGAAAACCCGCGCGAAAACCCGCGAAAACCCGCGAAAACCCGCGAAAACCCGGCGAAAACCCGGCGAAAACCCGGCGAAAACCCGGCGAAAACCCGGCGAAAACCCGGCGAAAACCCGGCGAAAACCCGGCGAAAACCCGGCGAAAACCCGGCGAAAAACCCGGCGAAAAACCGGCGAAAAACCGGCGAAAAACCGGCGAAAAACCGGCGAAAAAA	h H
1 GI 281 GACTTA 2 281 GACTTA 3 281 GACTTA 4 281 GACTTA 5 281 GACTTA 6 281 GACTTA 7 GII 281 GACTTA 8 281 GACTTA 10 281 GACTT 10 281 GACTT 11 281 GACTT 11 281 GACTT 12 281 GACTT 13 GII 281 GACTT 11 281 GACTT 13 GII 281 GACTT 11 281 GACTT 11 281 GACTT 11 281 GACTT 11 281 GACTT 12 281 GACTT 13 GII 281 GACTT 13 GII 281 GACTT 14 AGTT 15 AGTT 16 AGTT 17 GII 281 GACTT 18 GACTT 18 GACTT 18 GACTT 19 GACTT 19 GACTT 10 AGTT 10 AGTT 11 AGTT 12 AGTT 13 GIII 281 GACTT 14 GACTT 15 AGTT 16 AGTT 17 AGTT 18	GTCG TINICIGIGA GTCG TINICIGIGA GTCG TINICIGIGA GTCG TINICIGIGA GTCG TINICIGIGA GTCG TINICIGIGA GTCG TINICIGIGA GTCG TINICIGIGA GTCG TINICIGIGA GTCG TINICIGIGA	AAGCGCGGGG G AAGTGCGGGG G GAGTGCGGGG G AAGTCGGGGG G AAGCGCGGGG AAGCGCGGGA AAGCGCGCGGA AAGCGCGCGGA AAGCGCGCGGA AAGCGCGCGGA AAGCGCGCGGA AAGCGCGCGGA AAGCGCGCGGA AAGCGCGCGGA AAGCGCGGGA AAGCGCGCGGA AAGCGCGCGGA AAGCGCGGGA AAGCGCGCGGA AAGCGCGCGGA AAGCGCCGGGA AAGCGCGGGA AAGCGCCGGGA AAGCGCCCGGGA AAGCGCCGGGA AAGCGCCGCGGA AAGCGCCCGGA AAGCGCCCGGA AAGCGCCCGGA AAGCGCCCGGGA AAGCGCCCGGGA AAGCGCCCGGGA AAGCGCCCGGA AAGCGCCCCGGA AAGCGCCCGGA AAGCGCCCGGA AAGCGCCCGCGA AAGCGCCCGGA AAGCGCCCGGA AAGCGCCCGGA AAGCGCCCGGA AAGCGCCCGGA AAGCGCCCGGA AAGCGCCCGCGA AAGCGCCCGGA AAGCGCCCGCGA AAGCGCCCGGA AAGCGCCCGGA AAGCGCCCGGA AAGCGCCCGA AAGCGCCCGA AAGCGCCCGA AAGCCCCCGGA AAGCCCCCG	STCCAGGAGG STCCAGGAGG STCCAGGAGG STCCAGGAGG STCCAGGAGG ACCCAGGAGG ACCCAGGAGG ACCCAGGAGG ACCCAGGAGG ACCCAGGAGG ACCCAGGAGG	ACGCGGCGAG ACGCGCGCGAG ACGCGCGCGAA ACGCGGCGAA ACGCGGCGAA ACGCGGCAAAG ACGCGGCAAAG ACGCGGCAAAG ACGCGGCAAAG ACGCGGCAAAG ACGCGGCAAAG ACGCGGCAAAG ACGCGGCAAAG ACGCGGCAAAG ACGCGGCGAA	n u
2 281 GACTTA 4 281 GACTTA 5 281 GACTTA 6 281 GACTTA 6 281 GACTTA 7 GII 281 GACCTA 8 281 GACCTA 10 281 GACCTA 11 281 GACCTA 11 281 GACCTA 12 281 GACCTA 13 GACTTA 13 GISI 281 GACCTA	GTCG TIAICTGTGA GTCG TIAICTGTGA GTCG TIAICTGTGA GTCG TIAICTGTGA GTCG TIAICTGTGA GTCG TIAICTGTGA GTCG TIAICTGTGA GTCG TIAICTGTGA GTCG TIAICTGTGA GTCG TIAICTGTGA	AAGTGCGGGG G GAGTGCGGGG G GAGTGCGGGG G AAGTCGGGGG AAGTGCGGGG AAGTGCGGGG AAGTGCGGGG AAGTGCGGGG AAGTGCGGGGA GAGTGCGGGGA GAGTGCGGGA GAGTGCGGGGA GAGTGCGGGGA GAGTGCGGGGA GAGTGCGGGGA GAGTGCGGGGA GAGTGCGGGGA GAGTGCGGGA GAGTGCGGGA GAGTGCGGGGA GAGTGCGCGGGA GAGTGCGCGCGGA GAGTGCGCGGGA GAGTGCGCGGGA GAGTGCGCGGGA GAGTGCGCGGA GAGTGCGCGGGA GAGTGCGCGGGA GAGTGCGCGGGA GAGTGCGCGGGA GAGTGCGCGGA GAGTGCGCGGGA GAGTGCGCGGGA GAGTGCGCGGGA GAGTGCGCGGGA GAGTGCGCGGA GAGTGCGCGGGA GAGTGCGCGGA GAGTGCGCGGA GAGTGCGCGGA GAGTGCGCGGA GAGTGCGCGGGA GAGTGCGCGGA GAGTGCGCGGA GAGTGCGCGGA GAGTGCGCGGA GAGTGCGCGGGA GAGTGCGCGGA GAGTGCCGCGGA GAGTGCCGCGGA GAGTGCCGCGGA GAGTGCCGCGGA GAGTGCCGCGGA GAGTGCCCGGA GAGTGCCGCGGA GAGTGCCCGCGA GAGTGCCCGCGA GAGTGCCCGCGA GAGTGCCCGGA GAGTGCCCGGA GAGTGCCCGGA GAGTGCCCCGGA GAGTGCCCCGGA GAGTGCCCCGGA GAGTGCCCCCGGA AAGTGCCCCCGGA AAGTGCCCCCGGA AAGTGCCCCCGGA AAGTGCCCCCGGA AAGTGCCCCCGGA AAGTGCCCCCGGA AAGTGCCCCCGGA AAGTGCCCCCGGA AAGTGCCCCCCCCCA AAGTGCCCCCCCCCA AAGTGCCCCCCCA AAGTGCCCCCCCA AAGTGCCCCCCA AAGTGCCCCCCA AAGTGCCCCCCA AAGTGCCCCCCA AAGTGCCCCCCA AAGTGCCCCCCA AAGTGCCCCCCA AAGTCCCCCCCA AAGTCCCCCCCA AAGTCCCCCCCA AAGTCCCCCCCA AACTCCCCCCCA AACTCCCCCCCA AACTCCCCCCA AACTCCCCCCA AA	STCCAGGAGG STCCAGGAGG STCCAGGAGG STCCAGGAGG STCCAGGAGG ACCCAGGAGG ACCCAGGAGG ACCCAGGAGG ACCCAGGAGG ACCCAGGAGG	ACGCGGCGAG ACGCGCGCGAA ACGCGGCGAA ATGCGGCGAA ACGCGGCAAAG ACGCGGCAAAG ACGCGGCAAAG ACGCGGCAAAG ACGCGGCAAAG ACGCGGCAAAG ACGCGGCAAAAAAAA	# 8
3 281 GACTTG 4 281 GACTTA 5 281 GACTTA 6 281 GACTTA 7 GII 281 GACCTI 8 281 GACCTI 9 281 GACCTI 10 281 GACCTI 11 281 GACCTI 11 281 GACCTI 12 281 GACCTI 13 GACCTI 14 281 GACCTI 16 281 GACCTI 17 GII 281 GACCTI 18 GACCTI 19 GACCTI 11 281 GACCTI 11 281 GACCTI 11 281 GACCTI 12 281 GACCTI 13 GACCTI 13 GACCTI 14 281 GACCTI 15 ACCTI 16 ACCTI 17 ACCTI 18 GACCTI 18 GACCTI 19 GACCTI 19 GACCTI 10 GACCTI 11 281 GACCTI 11 281 GACCTI 12 281 GACCTI 13 GACCTI 13 GACTTI 14 GACCTI 15 GACCTI 15 GACCTI 16 GACCTI 17 GACCTI 18 GACCTI	GOTCG TIATCTGTGA GOTCG TIATCTGTGA GOTCG TIATCTGTGA GOTCG TIATCTGTGA GOTCG TIATCTGTGA GOTCG TIATCTGTGA GOTCG TIATCTGTGA GOTCG TIATCTGTGA GOTCG TIATCTGTGA	GAGTGCGGGG GAGTGCGGGG GAGTGCGGGG GAGTGCGGGGG AAGCGCGGG AAGCGCGGGA AAGCGCGGGA AAGCGCGGGA AAGCGCGGGA AAGCGCGGGA GAGCGCGGGA GAGCGCGGA GAGCGCGGGA GAGCGCGGGA GAGCGCGGGA GAGCGCGGA GAGCGCGGA GAGCGCGGA GAGCGCGGGA GAGCGCGGGA GAGCGCGGA GAGCGCGGA GAGCGCGGA GAGCGCGGA GAGCGCGGGA GAGCGCGGA GAGCGCGCGGA GAGCGCGGA GAGCGCGCGGA GAGCGCGCGGA GAGCGCGCGA GAGCGCGGA GAGCGCGCGA GAGCA	STCCAGGAGG STCCAGGAGG STCCAGGAGG STCCAGGAGG AACCAAGAGG ACCCAGGAGG ACCCAGGAGG ACCCAGGAGG ACCCAAGAGG	ACGCGGCGAA ATGCAGCGAA ATGCGGCGAA ACGCGGCAAG ACGCGGCAAG ATGCGGCGAAG ACGCGGCGAAG	<u>n</u>
4 281 GACTTA 5 281 GACTTA 6 281 GACTTA 7 GII 281 GACCTI 8 281 GACCTI 9 281 GACCTI 10 281 GACCTI 11 281 GACCTI 11 281 GACCTI 12 281 GACCTI 13 GACTTI 13 GACTTI 14 281 GACCTI 15 281 GACCTI 16 281 GACCTI 17 GIII 281 GACCTI 18 GACCTI 19 GACCTI 10 GACCTI 11 281 GACCTI 11 281 GACCTI 12 281 GACCTI 13 GACTTI 14 GACCTI 15 CACTI 15 CACTI 16 CACTI 17 CACTI 18 CACTI 19 CACTI 19 CACTI 19 CACTI 19 CACTI 10 CACTI 11 CACT	GTCG TIATCTGTGA GTCG TIATCTGTGA GTCG TIATCTGTGG GTCG TIATCTGTGG GTCG TIATCTGTGG GTCG TIATCTGTGG GTCG TIATCTGTGG	AAGTGCGGGA (AAGTCGCGGGA (AAGCGCGGGA AAGCGCGGGA AAGCGCGGGA (AAGCGCGGGA AAGCGCGCGGA AAGCGCCGGGA AAGCGCGCGGA AAGCGCGCGGA AAGCGCCGGGA AAGCGCCGGGA AAGCGCCGGGA AAGCGCCGGGA AAGCGCCGCGGA AAGCGCCGGGA AAGCGCCCGGGA AAGCGCCCGGGA AAGCGCCCGGGA AAGCGCCCGGGA AAGCGCCCGGGA AAGCGCCGCGGA AAGCGCCGCGGA AAGCGCCCGGGA AAGCGCCCGGGA AAGCGCCCGGGA AAGCGCCCGGGA AAGCGCCCGGGA AAGCGCCCGGGA AAGCGCCCGGGA AAGCGCCCGGGA AAGCGCCCGGA AAGCGCCCGGA AAGCGCCCGGA AAGCGCCCGGA AAGCGCCCGGA AAGCGCCCGGGA AAGCGCCCGGA AAGCGCCCCGGA AAGCGCCCGGA AAGCGCCCGGA AAGCGCCCGGA AAGCGCCCCGGA AAGCGCCCGGA AAGCGC	STCCAGGAGG STCCAGGAGG ATCCAGGAGG ACCCAGGAGG ACCCAGGAGG ACCCAGGAGG ACCCAGGAGG ACCCAGGAGG	ACCCGCCGAA ATGCACCGAA ACCCGCCGAG ACCCGCCAAG ATGCGCCAAG ACCCGCCAAA ACCCGCCGAA	# #
5 281 GACTTA 6 281 GACCTA 7 GII 281 GACCTI 8 281 GACCTI 9 281 GACCTI 10 281 GACCTI 11 281 GACCTI 11 281 GACCTI 12 281 GACCTI 13 GII 281 GACCTI 13 GII 281 GACCTI 14 281 GACCTI 15 281 GACCTI 16 281 GACCTI 17 GII 281 GACCTI	(GTCG TIATCTGTG) 18 18 18 18 18 18 18 18 18 18 18 18 18 1	AAGTCAGGGA (AAGTGCGGGG (AAGCGCGGGA I	STCCAGGAGG STCCAGGAGG ACCCAGGAGG ACCCAGGAGG ACCCAGGAGG ACCCAGGAGG ACCCAAGAGG	ATGCAGCGAA ACGCGGCGAG ACGCGGCAAG ACGCGGCGAA ACGCGGCGAA ACGCGGCGAA	n u
6 281 GACCTA 7 GII 281 GACCTI 8 281 GACCTI 9 281 GACCTI 10 281 GACCTI 11 281 GACCTI 11 281 GACCTI 12 281 GACCTI 13 GIII 281 GACCTI 13 GIII 281 GACCTI	GTCG TIATCTGCG) GTCG TIATCTGTGG GTCG TIATCTGTGG GTCG TIATCTGTGG GTCG TIATCTGTGG GTCG TIATCTGTGG GTCG TIATCTGTGGG GTCG TIATCTGTGGGG GTCG TIATCTGTGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	AAGCGCGGGA AAGCGGCGGA AAGCGGCGGA AAGCGGCGGA AAGCGGCGGA AAGCGGCGGA AAGCGGCGGGA AAGCGGCGGA AAGCGGCGGGA AAGCGCGGGA AAGCGGCGGGA AAGCGGCGGGA AAGCGGCGGGA AAGCGGCGGGA AAGCGGCGGGA AAGCGGCGGGA AAGCGGCGGA AAGCGGCGGGA AAGCGGCGGA AAGCGGA AAGCGGCGGA AAGCGGCGGA AAGCGGCGGA AAGCGGCGGA AAGCGGCGGA AAGCGGCG	STCCAGGAGG BEREINE ENTER AACCAGGAGG ACCCAGGAGG ACCCAAGAGG ACCCAAGAGG	ACGCGGCGAG BERRERERE ACGCGGCAAG ATGCGGCGAA ACGCGGCGAA ACGCGGCGAA	n u
1 GII 281 GACCTI 8 281 GACCTI 9 281 GACCTI 10 281 GACCTI 11 281 GACCTI 12 281 GACCTI 13 GACCTI 14 281 GACCTI 15 281 GACCTI 16 281 GACCTI 17 GIEI 281 GACCTI 18 GACTTI	GCTCG TTATCTGTGG GCTCG TTATCTGTGG GCTCG TTATCTGTGG GCTCG TTATCTGCGG GCTCG TTATCTGTGG	AAGCGGGGA AAGCGCGGGA AAGCGCGGGA GAACGCGGGAA GAACGCGGGGAA GAACGCGCGGGAA GAACGAAC	AACCAGGAGG ACCCAGGAGG ACCCAGGAGG ACCCAGGAGG	ACGCGGCAAG ATGCGGCGAG ATGCGGCGAA ACGCGGCGAA ACGCGGCGAG	m u
n 0		AAGCGCGGGA AAGCGCGGGA GAGCGCGGGA GAGCGCGGGA AGCGCGGGA AGCGCGGGGA AGCGCGGGGA AGCGCGGGA AGCGCGGGGA AGCGCGGGA AGCGCGGGGA AGCGCGGGA AGCGCGGGGA AGCGGGGGA AGCGGGGGA AGCGGGGGA AGCGCGGGGA AGCGCGGGGA AGCGCGGGGA AGCGCGGGGA AGCGGCGGGA AGCGGCGGGA AGCGGCGGGA AGCGGCGGGA AGCGGCGGGA AGCGGCGGGA AGCGGCGGGA AGCGGGGA AGCGGCGGGA AGCGGGGGA AGCGGGGGA AGCGGGGGA AGCGGGGGA AGCGGGGGA AGCGGCGGGA AGCGGGGGA AGCGGGGGA AGCGGGGGA AGCGGGGGA AGCGGGGGA AGCGGGGA AGCGGGGGA AGCGGGGGA AGCGGGGGA AGCGGGGGA AGCGGGGGA AGCGGGGGA AGCGGGGGA AGCGGGGGA AGCGGGGGA AGCGGGGGGA AGCGGGGGGA AGCGGGGGGA AGCGGGGGGA AGCGGGGGGA AGCGGGGGGA AGCGGCGGGGA AGCGGCGGGGGA AGCGGCGGGGA AGCGGCGGGGGA AGCGGCGGGGGA AGCGGCGGGGGGA AGCGGCGGGGGGA AGCGGGGGGGA AGCGGGGGGGA AGCGGGGGGGA AGCGGGGGGGG	ACCCAGGAGG ACCCAGGAGG ACCCAAGAGG	ATGCGGCGAG ACGCGGCGAA ACGCGGCGAG ACGCGGCGAG	
n 0		AAGCGCGGGA AAGCGGGA AAGCGCGGGA AAGCGCGGGA AAGCGCGGGA AAGCGCGGGA AAGCGCGGGA AAGCGGA AAGCGCGGGA AAGCGCGGA AAGCGCGGA AAGCGCGGA AAGCGGCGGA AAGCGGCGGA AAGCGGA AAGCGCGGA AAGCGGCGGA AAGCGGCGGA AAGCGGCGGA AAGCGGCGGA AAGCGCGGA AAGCGGCGGA AAGCGCGGA AAGCGGCGGA AAGCGCGGA AAGCGCGGA AAGCGCGA AAGCGCGGA AAGCGCGA AAGCGGCGA AAGCGCGA AAGCGCA AAGCGCA AAGCGCA AAGCA AAGCGCA AAGCGA AAGCGA AAGCGA AAGCA AA	ACCCAGGAGG ACCCAGGAGG ACCCAAGAGG	Argeogedaa Aegeogedaa Aegeogedaa	
n 0		AAGCGCGGGA GAGCGCGGGA GAGCGCGGGA	ACCCAGGAGG ACCCAAGAGG ACCCAAGAGG	ACGCGGCGAG ACGCGGCGAG ACGCGGCGAG	
n 0		GAGCGCGGGA	ACCCAAGAGG	ACGCGGCGAG	
n 0		GAGCGCGGGA	ACCCAAGAGG	ACGCGGCGAG	
n n		できせていたいでんで	1		
n 0		י פאסטטטטטטטטט	ACCCAAGAGG	TTAICIGIGA GAGCGCGGGG ACCCAAGAGG ACGCGGCGAG	CCLACGAGIC
281	010000000000000000000000000000000000000	- 10 年 10			
		AAGCCAGGGG	ACTGAGGAGG	ACGAGCGGAA	
14 281 GACCIGGICG	3GTCG TCATCTCAG			ACGAGCAGAA	
15 281 GACCIGGTIG	SGTTG TCATCTCAGA	GAGTCAGGGG		ATGAGCGGAA	
16 281 GACCTAGTCG		TCATCTCAGA "GAGTCAAGGG	GTCGAGGAGG	ATGAGCGAAA	
281	GACCTGGTCG TCATCTCGG	TCATCTCGGA GAGCGAAGGT	AACGAGGAGG	ACGAGCGAAA	CCTGAGAGCT
	BORGORMENSESSESSESSESSESSESSESSESSESSESSESSESSE				
	מופר בנאודות	000000000000000000000000000000000000000	000000000000000000000000000000000000000	りについていたして	
	ACCITGGIGG CCATITICGA GAGCCAAGGG ACGCACGAGG ATGAAGCGIG CCIGAGAGIC	A GAGCCAAGGG ACGCACGAGG	ACGCACGAGG	ATGAAGCGTG	CCIGAGAGIC
20 GIV 281 GATCT	DATCIBOTO TOTIOCION DAGIONICOC GICGACGAGO ATAGAGCAGC	A GAGTGATGGC	GTCGACGAGG	ATAGAGCAGC	CCTGAGAGCC
281	GAICIGGIIG IGGIGGCIGA GAGIGAIGGC GICGACGAGG AIAGAACAGC CCIGCGAGCC	A GAGTGATGGC	GICGACGAGG	ATAGAACAGC	CCTGCGAGCC
281	GATCTGGTTG TGGTGGCTGA GAGTGATGGC GTCAATGAGG ATAGAGCAGC	A GAGTGATGGC	GTCAATGAGG	ATAGAGCAGC	CCTGGGAGCC

340 TOTAL

7/2/

Fig. 3

ENVELOPE REGION

ID NUMBER	Genotype					٠		
2 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5	# # # # # # # # # # # # # # # # # # #	ÿ # # → → →	GACGGCGTTG GACGGCGTTG AACGGCGCTG	# 60 C	AGCTGCTCCG AGGTACTCCG AGCTGCTCAG	GATCCCACAA GATCCCACAA GATCCCACAA GGTCCCGCAA	GCCATCATGG GCCATCATGG GCCATCGTGG	GENERON BENERAL RESTERENCE BENER A ACATGATCGC ACATGATCGC ACATGATCGC
26 27 28	GII	8 8 0 0 0	GOCAGCCTA GOCAGCCTA GOCAGCCTA	GIGGIGICGC GIGGIGICGC GIGGIGICGC	AGTTACTCCG AGTTACTCCG AGTTACTCCG	AGTIACTCCG GATCCCACAA AGTIACTCCG GATCCCACAA AGTIACTCCG GATCCCGCAA	GCCGTCATGG AGCATCGTGG GCTGTCGTGG	ATATGGTGGC ACATGGTGGC ACATGGTGGC
30 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	62 GIV GIV 31	8 8 8 8 4 4 4 1	TGTGGGTATG TGTGGGTATG TGTGGGTATG	IGIGGGIAIG GIGGIGGCGC ACGICCIGCG TITGCCCCAG ACCTIGITCG ACAIAATAGC TGIGGGIAIG GIGGIAGCAC ACGICCIGCG ICTGCCCCAG ACCTIGITCG ACAIAATAGC IGIGGGIAIG GIGGIGGCG AAGICCIGCG TITGCCCCAG ACCTIGITCG ACGIGCTAGC	ACGICCIGCG ACGICCIGCG AAGICCIGCG	TTTGCCCCAG TTTGCCCCAG	ITIGCCCCAG ACCTIGITCG ICIGCCCCAG ACCTIGITCG ITIGCCCCCAG ACCTIGITCG	ACATAATAGC ACATAATAGC ACGTGCTAGC
32 11 11 11 11 11 11 11 11 11 11 11 11 11	32 GIII			CTCCTGGCAT ACTTGGTGCG	CCTGGCAT ACTTGGTGCG CATCCCGGAG	CATCCCGGAG	CATCCGGAG GTCATCCTGG .	ACATTATCAC
22 23 23 23 23 23 23 23 23 23 23 23 23 2	15		TGGTGCTCAC TGGAGCCCAC TGGTGCCCAC	TGGGGAGTCC TGGGGAGTCC	TGGCGGGCAT TGGCGGGCAT :TAGCGGGCAT	AGCGTATITIC AGCGTATITIC AGCGTATITIT		
26 26 27 27 28	26 GII	61 61 61 61	#	GGGGGCCCAC TGGGGAGTCC TGGCGGGCCT TGCCTACTAT GGGGGCCCAC TGGGGAGTCC TGGCGGGCCT TGCTTACTAT GGGGGCCCAC TGGGAATCC TAGCGGGTCT TGCCTACTAT	TGGCGGGCCT TAGCGGGCCT TAGCGGGTCT	TGCCTACTAT TGCTTACTAT TGCCTACTAT		
31 31	29 GIV 30 31	61 61 61 61 61 61 61 61 61 61 61 61 61 6	CGGGGCCCAT	I TGGGGCATCT I TGGGGCATCT I TGGGGCATCT	TOGGGCATCT TGGCGGGCTT TGGGGCATCT TGGCGGGCCT TGGGGCATCT TGGCGGGCCT	GGCCTATTAC AGCCTATTAC GGCCTATTAC		·
88558888 32	annonnuncanuncung 32 GIII	10000 61	Bergergerete GGGAGGACAC		TOGGGCGICA TOTTTGGCCT GGCTTATTTC	. OGCITATITC	11 12 13	

100 Total

Fig. 4

D NUMBER GENOTYPE	SEQUENCE								
33 01 1 0TTAGTATGA GYGTCGGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 34 1 0TTAGTATGA GYGTCGGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 35 1 0TTAGTATGA GYGTCGTGCA GYCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 35 1 0TTAGTATGA GYGTCGTGCA GYCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 37 1 0TTAGTATGA GYGTCGTGCA GYCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 40 1 0TTAGTATGA GYGTCGTGCA GYCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 41 1 0TTAGTATGA GYGTCGTGCA GYCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 42 1 0TTAGTATGA GYGTCGTGCA GYCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 43 1 0TTAGTATGA GYGTCGTGCA GYCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 43 1 0TTAGTATGA GYGTCGTGCA GYCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 45 1 0TTAGTATGA GYGTCGTGCA GYCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCC 45 0TTAGTATGA GYGTCGTGCA GYCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCC 45 0TTAGTATGA GYGTCGTGCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCC 45 0TTAGTATGA GYGTCGTGCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCC 45 0TTAGTATGA GYGTCGAGAGC CCCCCCTCC CGGGAGAGCC ATAGTGGTC 45 0TTAGTAGA GYCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGG	ID NUMBER	GENOTYPE							
1 GTTAGTATGA GTGTCCAGGA CCCCCCTCC CGGGAGGCC ATAGTGGTCT 35 1 GTTAGTATGA GTGTCGTGCA GCCCCCTCC CGGGAGGCC ATAGTGGTCT 35 1 GTTAGTATGA GTGTCGTGCA GCCCCCCTCC CGGGAGGCC ATAGTGGTCT 3 GTTAGTATGA GTGTCGTGCA GCCTCCAGGA CCCCCCCTCC CGGGAGGCC ATAGTGGTCT 40 GTTAGTATGA GTGTCGTGCA GCCTCCAGGA CCCCCCCTCC CGGGAGGCC ATAGTGGTCT 41 GTTAGTATGA GTGTCGTGCA GCCTCCAGGA CCCCCCCTCC CGGGAGGCC ATAGTGGTCT 41 GTTAGTATGA GTGTCGTGCA GCCTCCAGGA CCCCCCTCC CGGGAGGCC ATAGTGGTCT 42 GTTAGTATGA GTGTCGTGCA GCCTCCAGGA CCCCCCTCC CGGGAGGCC ATAGTGGTCT 44 GTTAGTATGA GTGTCGTGCA GCCTCCAGGA CCCCCCCTCC CGGGAGGCC ATAGTGGTCT 45 GTTAGTATGA GTGTCGTGCA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 45 GTTAGTATGA GTGTCGTGCA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 45 GTTAGTATGA GTGTCGTGCA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 45 GTTAGTATGA GTGTCGTACA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 45 GTTAGTATGA GTGTCGTACA CCCCCTCC CGGGAGAGCC ATAGTGGTCT 45 GTTAGTATGA GTCTCCAGGA CCCCCTCC CGGGAGAGCC ATAGTGGTCT 45 GTTAGTATGA GTGTCGTAGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 45 GTTAGTAGTAGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT GTGTCGTAGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 45 GTTAGTGTCT CGGGAGAGCC CGGGAGAGCC ATAGTGGTCT 45 GTTAGTAGTCT CGGGAGAGCC C		unannanan ID	8 C	GITAGIATGA	GTGTCGTGCA	GCCTCCAGGA	CCCCCCTCC	CGGGAGAGCC	ATAGTGGTCT
	34		-	GTTAGTATGA			CCCCCCTCC	CGGGAGAGCC	ATAGTGGTCT
	35		-	GTTAGTATGA	GTGTCGTGCA		CCCCCCTCC		ATAGTGGTCT
17 1 GTTAGTATGA GTGTCGGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 18 GTTAGTATGA GTGTCGTGCA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 19 GTTAGTATGA GTGTCGTGCA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 41 1 GTTAGTATGA GTGTCGTGCA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 43 1 GTTAGTATGA GTGTCGTGCA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 44 1 GTTAGTATGA GTGTCGTGCA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 45 1 GTTAGTATGA GTGTCGTGCA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 45 1 GTTAGTATGA GTGTCGTGCA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 45 1 GTTAGTATGA GTGTCGTGCA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 46 11 1 GTTAGTATGA GTGTCGTGCA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 47 1 GTTAGTATGA GTGTCGTGCA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 48 GIN 1 GTTAGTATGA GTGTCGTACA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 48 GIN 1 GTTAGTATGA GTGTCGTACA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 48 GIV 1 GTTAGTATGA GTGTCGTACA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 49 GTTAGTATGA GTGTCGTACA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 40 GTTAGTATGA GTGTCGTACA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 40 GTTAGTATGA GTGTCGTACA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 41 GTTAGTATGA GTGTCGTACA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 42 GTTAGTATGA GTGTCGTACA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 43 GTTAGTATGA GTGTCGTACA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 44 GIV 1 GTTAGTATGA GTGTCGTACA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 45 GTTAGTATGA GTGTCGTACA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 46 GIV 1 GTTAGTATGA GTGTCGTACA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 51 GTTAGTATGA GTGTCGTACA GCCTCCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 51 GTTAGTATGA GTGTCGTACA GCCTCCC CGGGAGAGCC ATAGTGGTCT 51 GTTAGTATGA GTGTCGTACA GCCTCCCCTCC CGGGAGAGCC ATAGTGGTCT 51 GTTAGTATGA GTGTCGTACA GCCTCCCCCCCCCCCCC	36		-	GTTAGTATGA	GIGICGIGCA	GCCTCCAGGA		CGGGAGAGCC	ATAGTGGTCT
	37		-	OTTAGTATGA			CCCCCCTCC	CGGGAGAGCC	ATAGTGGTCT
0	38			GTTAGTATGA	GIGICGIGCA	GCCTCCAGGA	ودودودوعود	CGGGAGAGCC	ATAGIGGICI
40 1 GTTAGTATGA GTGTCGTGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 41 1 GTTAGTATGA GTGTCGTGCA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 43 1 GTTAGTATGA GTGTCGTGCA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 44 1 GTTAGTATGA GTGTCGTGCA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 45 1 GTTAGTATGA GTGTCGTGCA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 46 GII		IID	8	GTTAGTATGA	GTGTCGTGCA	GCCTCCAGGA		CGGGAGAGCC	ansmenen ATAGTGGTCT
41 1 OTTAGTATGA GTGTCGAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 43 1 GTTAGTATGA GTGTCGTGCA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 44 1 GTTAGTATGA GTGTCGTGCA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 45 1 GTTAGTATGA GTGTCGTGCA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 45 1 GTTAGTATGA GTGTCGTGCA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 46 GII 1 GTTAGTATGA GTGTCGTGCA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 47 1 GTTAGTATCA GTGTCGTACA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 48 GIV 1 GTTAGTATGA GTGTCGTACA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 49 GTTAGTATGA GTGTCGTACA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 49 GTTAGTATGA GTGTCGTACA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 49 GTTAGTATGA GTGTCGTACA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 40 GTTAGTATGA GTGTCGTACA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 40 GTTAGTATGA GTGTCGTACA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 40 GTTAGTATGA GTGTCGTACA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 50 GV 1 GTTAGTATGA GTGTCGTACA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 51 GTTAGTATGA GTGTCGTACA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 51 GTTAGTATGA GTGTCGTACA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 51 GTTAGTATGA GTGTCGTACA GCCTCCAGGA CCCCCTCC CGGGAGAGCC ATAGTGGTCT 51 GTTAGTATGA GTGTCGTACA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 51 GTTAGTATGA GTGTCGAACA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 51 GTTAGTATGA GTGTCGAACA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 51 GTTAGTAGTATGA GTGTCGAACAC CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 51 GTTAGTAGTATGA GTGTCGAACAC CCCCCCTCC CGGGAGAGCC ATAGTGGTCTCC CGGGAGAGCC ATAGTGGTCTCC CGGGAGAGCC ATAGTGGTCTCC CGGGAGACC ATAGTGGTCTCC CGGGAGAGCC ATAGTGGTCTCC CGGGAGACC ATAGTGC	9	•	-	GTTAGTATGA	GTGTCGTGCA	GCCTCCAGGA	CCCCCCTCC	CGGGAGAGCC	ATAGIGGICI
42 1 GTTAGTATGA GTGTCGTGCA GCCCCCCTCC CGGGAGAGCC ATAGTGGTCT 44 1 1 GTTAGTATGA GTGTCGTGCA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 45 1 GTTAGTATGA GTGTCGTGCA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 46 1 GTTAGTATGA GTGTCGTGCA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 46 GIII 1 GCTAGTATGA GTGTCGTACA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 47 1 GCTAGTATCA GTGTCGTACA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 48 GIV 1 GTTAGTATGA GTGTCGTACA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 48 GIV 1 GTTAGTATGA GTGTCGTACA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 49 GTTAGTATGA GTGTCGTACA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 49 GTTAGTATGA GTGTCGTACA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 49 GTTAGTATGA GTGTCGTACA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 40 GTTAGTATGA GTGTCGTACA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 50 GV 1 GTTAGTATGA GTGTCGAACA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 51 GTTAGTAGTAGA GTGTCGAACA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCTCC CGGGAGAGCC ATAGTGGTCT CCCCCTCC CGGGAGAGCC ATAGTGGTCT CCCCCCTCC CGGGAGAGCC ATAGTGGTCT CCCCCCTCC CGGGAGAGCC ATAGTGGTCT CCCCCCTCC CGGGAGAGCC ATAGTGGTCT CCCCCCTCC CGGGAGAGCC ATAGTGGTCT CCCCCCTCCCCCCCCCC	41		-	GTTAGTATGA	_	GCCTCCAGGA			ATAGTGGTCT
43 1 GTTAGTATGA GTGTCGTGCA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 44 1 GTTAGTATGA GTGTCGTGCA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT ===============================	42	•	-	GITAGIATGA	GTGTCGTGCA			CGGGAGAGCC	ATAGTGGTCT
44 1 GTTAGTATOA GTGTCGTGCA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT ===============================	43		~	GTTAGTATGA		GCCTCCAGGA		-	ATAGTGGTCT
45 1 GTTAGTATGA GTGTCGTGCA GCCTCCAGGA CCCCCCTCC CGGGAGGCC ATAGTGGTCT 46 GIII 1 GCTAGTATCA GTGTCGTACA GCCTCCAGGC CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 47 1 GTTAGTATGA GTCTCGTACA GCCTCCAGGC CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 48 GIV 1 GTTAGTACA GTGTCGTACA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 49 1 GTTAGTACA GTGTCGTGCA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 49 1 GTTAGTACA GTGTCGTGCA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 49 1 GTTAGTACA GTGTCGTGCA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 50 GV 1 GTTAGTATGA GTGTCGTACA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 51 1 GTTAGTATGA GTGTCGTACA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 51 1 GTTAGTATGA GTGTCGTACA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 51 1 GTTAGTATGA GTGTCGAACA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 51 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	44		-	GTTAGTATGA	-	GCCTCCAGGA	CCCCCCTCC		ATAGTGGTCT
46 GIII 1 GCTAGTATCA GTĞTCGTACA GCCTCCAGGC CCCCCCTCC ÇGGGAGAGCC ATAGTGGTCT 47 1 GTTAGTATGA GTCTCGTACA GCCTCCAGGC CCCCCCTCC ÇGGGAGAGCC ATAGTGGTCT 48 GIV 1 GTTAGTACGA GTGTCGTGCA GCCTCCAGGA CTCCCCCTCC CGGGAGAGCC ATAGTGGTCT 49 1 GTTAGTACGA GTGTCGTGCA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 49 1 GTTAGTATGA GTGTCGTGCA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 50 GV 1 GTTAGTATGA GTGTCGAACA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 51 1 GTTAGTATGA GTGTCGAACA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT	45		-	GITAGIAIGA	GIGICGIGCA	GCCTCCAGGA	CCCCCCTCC	CGGGAGAGCC	ATAGIGGICI
47 1 GTTAGTATGA GTCTCGTACA GCCTCCAGGC CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 48 GIV 1 GTTAGTACGA GTGTCGTGCA GCCTCCAGGA CTCCCCCTCC CGGGAGAGCC ATAGTGGTCT 49 1 GTTAGTACGA GTGTCGTGCA GCCTCCAGGA CTCCCCCTCC CGGGAGAGCC ATAGTGGTCT 49 1 GTTAGTATGA GTGTCGTGCA GCCTCCAGGA CTCCCCCTCC CGGGAGAGCC ATAGTGGTCT 50 GV 1 GTTAGTATGA GTGTCGAACA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 51 1 GTTAGTATGA GTGTCGAACA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 51 1 GTTAGTATGA GTGTCGAACA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 51 1 GTTAGTATGA GTGTCGAACA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT		IIID	n 0 0 0 0 0 0 0	GCTAGTATCA	GTĞTCGTACA	GCCTCCAGGC		COGGAGAGCC	ATAGIGGICI
48 GIV 1 GTTAGTACGA GTGTCGTGCA GCCTCCAGGA CTCCCCCTCC CGGAGAGCC ATAGTGGTCTCT 49 1 GTTAGTACGA GTGTCGTGCA GCCTCCAGGA CCCCCCTCC CGGAGAGCC ATAGTGGTCTCT 50 GV 1 GTTAGTATGA GTGTCGAACA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCTCT 51 GTTAGTATGA GTGTCGAACA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT 51 GTTAGTATGA GTGTCGAACA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT	47		-	GTTAGTATGA	GICICGIACA	GCCTCCAGGC	CCCCCCTCC	CGGGAGAGCC	ATAGTGGTCT
49 1 GTTAGTACGA GTGTCGTGCA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT *******************************	5 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	AID		GTTAGTACGA	GIGICORGEA	GCCTCCAGGA	CTCCCCCTCC	COCOACACC	ATAGTGGTCT
HINDER HINDER HOLD TO GITAGIATE GIGICOLUNG CONTROL CON	9		-	GITAGIACGA	GTGTCGTGCA	GCCTCCAGGA	ככככככבבכ	CGGGAGAGCC	ATAGTGGTCT
51 1 GITAGIAIGA GIGICGAACA GCCICCAGGA CCCCCCICC CGGGAGAGCC AIAGIGGICI				GTTAGTATGA	GTGTCGAACA	GCCTCCAGGA		CGGGAGAGCC	ATAGTGGTCT
	51		-	GTTAGTATGA	GTGTCGAACA	GCCTCCAGGA	2012222222	CGGGAGAGCC	ATAGTGGTCT

Fig. 4b

11 11 12 13 14 15 15 15 15 15 15 15 15 15 15 15 15 15	AACCC	AACCC	GGATCAACCC	GGATAAACCC	GGATCAACCC	GGATAAACCC	GGATCAACCC	GGATCAACCC	GGATCAACCC	GGATCAACCC	00044	GGATCAACCC	GGATCAACCC	GGATCAACCC	11 12 13 13	AAACCC	GGATAAACCC		CAACCC	TAACCC	11 11 11 11 11 11 11 11 11 11 11 11 11	AAACCC	AAACCC	
0 0 11	GGATCAACCC	GGATCAACCC	GGATC	GGATA	GGATC	GGATA	GGATC	GGATC	GGATC	GGATC		1 6 6 6 6		GGAT	00000000000000000000000000000000000000	GGAT	GGAT		GGAG	CCAC	11 12 12 12 12 12 12 12 12 12 12 12 12 1	GGAT	GGAT	
33 35 30 90 90 91 91	recriterr		GICCLLICIT	GICCITICIT	GICCLIFICIT	GICCTIFCIT	Greenigh			CHALLET			greenterr	GICCILICII	. # 2 0 0 0 0 0 0 0 0 1	GICCILICIT	GICCITICIT		greerret	GICCLLINCIL	90000000000000000000000000000000000000	GICCITICIT	GICCLITCIL	
8 9 0 0 0 0 0 0 0	3GACGACCGG (SGACGACCGG		GGACGACCGG	GGACGACCGG	GGACGACCGG	GGACGACCGG	GGACGACCGG GTCCTTTCTT	GGACGACCGG	מטעמאנומט	000000000000000000000000000000000000000	GGACGACCGG	GGACGACCGG	GGACGACCGG		GGAAGACTGG	CGAAGACTOG		GGGTGACCGG	GGGTGACCGG		GGATGACCGG	GGATGACCGG	
17 10 11 13 14 14 14 14	GCGGAACCGG TGAGTACACC GGAATTGCCA GGACGACCGG GTCCTTTCTT	TOAGTACACC GGAATTGCCA GGACGACCGG	GGAATTGCCA (GGAATTGCCA (GGAATTGCCA GGACGACCGG	GCGGAACCGG IGAGIACACC GGAATIGCCA GGACGACCGG GICCIIICII	GGAATTGCCA GGACGACCGG GTCCTTTCTT	GGAATTRCCA	GGAATTGCCA		4339114400	TGAGTACACC GGAATTGCCA	TGAGTACACC GGAATTGCCA GGACGACCGG	GCGDAACCGG TGAGTACACC GGAATTGCCA GGACGACCGG		GEGGAACCGG TGAGTACACC GGAATTGCCG GGAAGACTGG GTCCTTTCTT GGATAAACCC	GCGGAACCGG TGAGTACACC GGAATTGCTG GGAAGACTGG GTCCTTTCTT		GCGGAACCGG TOAGTACACC GGAATCGCTG GGGTGACCGG GTCCTTTCTT GGAGCAACCC	GCGGAACCGG TGAGTACACC GGAATCGCTG GGGTGACCGG GTCCTTTCTT GGAGTAACCC		GCGGAACCGG TGAGTACACC GGAATTGCCG GGATGACCGG GTCCTTTCTT GGATAAACCC	GCGGAACCGG TGAGIACACC GGAATIGCCG GGAIGACCGG GICCTIICII GGAIAAACCC	
11 60 77 12 13 13 13 13 13 14 15 16 16 16 16 16 16 16 16 16 16 16 16 16	TGAGTACACC (TOAGTACACC (TGAGTACACC (TGAGTACACC (TGAGTACACC (TGAGTACACC	TGAGTACACC		TONGTHACKE CONTINCED TONGTHE	2010110101	TUARIBLACE	TGAGTACACC	TGAGTACACC	TGAGTACACC		TOAGTACACC	TGAGTACACC		TOAGTACACC	TGAGTACACC		TGAGTACACC	TGAGTACACC	
EQUENCE D NUMBER GENOTYPE	GCGGAACCGG	GCGGAACCGG	GCGGAACCGG		GCGGAACCGG	GCGGAACCGG	**************************************		DOUGHACTED	りついてもつうつう	GCGGAACCGG	GCGGAACCGG	GCGGAACCGG	GCGGAACCGG	HUBBUHAND DADA BOOGGESE MODER DE STEINE STEINE BOOK DE STEINE BOOK	GCGGAACCGG	GCGGAACCGG	自创的有效的自然的复数 医电阻性电子 计自己的 计通过电话 医克里特氏征 医克里特氏征 医克里特氏征 医克里特氏征 医克里氏征 医原生氏征 医原生氏征 医克里氏征 医原生 医原生氏征 医原生氏征 医原生氏征 医原生氏征 医原生氏征 医原生氏征 医原生氏征 医原生氏征 医原生氏征 医克里氏征 医原生氏征 医原生 医原生性 医原生性 医原生性 医原生性 医原生性 医原生性 医原生性	GCGGAACCGG	GCGGAACCGG		GCGGAACCGG	GCGGAACCGG	
	61	61	61	19	61	61	8 0 0 0 0 0 0 0 0 0	; ;	10	70	61	61	61	61		61	61		61	61		61	19	0 11 11 11 11
genotype	10							115								1110				;		20	•	
EQUENCE D NUMBER		34	. E.	9 6	37	38		٦ ·	0,5	1	45	43	44	. 4. R.			2 7	È	# # # # # # # # # # # # # # # # # # #	4 4			51	10 10 10 10 10 11 11

1421

Fid. 4

5'UT Region (3/5)

	AGTOTTGGGT AGTOTTGGGT AGTOTTGGGT AGTOTTGGGT AGTOTTGGGT AGTOTTGGGT AGTOTTGGGT	AGTGTTGGGT AGTGTTGGGT AGTGTTGGGT AGTGTTGGGT AGTGTTGGGT AGTGTTGGGT AGTGTTGGGT AGTGTTGGGT	AGCGTTGGGT AGCGTTGGGT AGTGTTGGGT AGTGTTGGGT	AGTOTTGGGT AGTOTTGGGT AGTOTTGGGT
	CTAGCCGAGT CTAGCCGAGT CTAGCCGAGT CTAGCCGAGT CTAGCCGAGT CTAGCCGAGT	SCICAATGCC TGGAGATITG GGCGTGCCCC CGCGAGACTG CTAGCCGAGT AGTGTTGGGT SCICAATGCC TGGAGATITG GGCGTGCCCC CGCGAGACTG CTAGCCGAGT AGTGTTGGGT GCTCAATGCC TGGAGATITG GGCGTGCCCC CGCGAGACTG CTAGCCGAGT AGTGTTGGGT	46 GIII 121 ACTCTATGCC CGGCCATTTG GGCGTGCCCC CGCAAGACTG CTAGCCGAGT AGCGTTGGGT 47 121 ACTCTATGCC CAGCCATTTG GGCGTGCCCC CGCAAGACTG CTAGCCGAGT AGCGTTGGGT 47 121 ACTCTATGCC CAGCCATTTG GGCGTGCCCC CGCAAGACTG CTAGCCGAGT AGCGTTGGGT 23 48 GIV 121 GCTCAATACC CAGAATTTG GGCGTGCCCC CGCGAGATCA CTAGCCGAGT AGTGTTGGGT	49 6121 CLICANIACE CASAGNITIS SOCIECES CONTROL CASAGNITIS SOCIECES CONTROL CASAGNITIS SOCIECES CASAGNITIS SOCIETES CASAGNITIS SOCIETES CASAGNITIS SOCIETES CASAGNITIS CASAGNITICA CASAGNITIS CASAGNITIS CASAGNITIS CASAGNITIS CASAGNITICA CA
	GGCGTGCCC CGCAAGACTG GGCGTGCCCC CGCAAGACTG GGCGTGCCCC CGCAAGACTG GGCGTGCCCC CGCAAGACTG GGCGTGCCCC CGCAAGACTG GGCGTGCCCC CGCAAGACTG	GCTCAATGCC TGGAGATITG GGCGTGCCCC CGCGAGACTG	CGCAAGACTG CGCAAGACTG CGCAAGACTG CGCGAGATCA	CCCCACACTCA CCCCACACTC CCCCACACTC
	GGCGTGCCC CGCAAGACTG GGCGTGCCCC CGCAAGACTG GGCACGCCCC CGCAAGACTG GGCGTGCCCC CGCAAGACTG GGCGTGCCCC CGCAAGACTG GGCGTGCCCC CGCAAGACTG	GGCGTGCCCC CGCGAGACTG GGCGTGCCCC CGCGAGACTG GGCGTGCCCC CGCGAGACTG GGCGTGCCCC CGCGAGACTG GGCGTGCCCC CGCGAGACTG		
	GCTCAATGCC TGGAGATTTG	GCTCAATGCC TGGAGATTTG GCTCAATGCC TGGAGATTTG GCTCAATGCC TGGAGATTTG GCTCAATGCC TGGAGATTTG GCTCAATGCC TGGAGATTTG GCTCAATGCC TGGAGATTTG	COGCCATITG CAGCCATITG CAGCAAATITG	CGGAGATTTG CGGAGATTTG
	GCTCAATGCC GCTCAATGCC GCTCAATGCC GCTCAATGCC GCTCAATGCC	GCTCAATGCC GCTCAATGCC GCTCAATGCC GCTCAATGCC GCTCAATGCC	ACTUTATEC ACTUTATEC ACTUTATEC GCTCAATACC	GCTCAATGCC GCTCAATGCC
	121 121 121 121 121 121	121 121 121 121 121 121 121 121 121 121	121 n n 1771 n n n n n n n n n n n n n n n	121 121 121 121
GENOTYPE	19 19	110		
SEQUENCE ID NUNBER	2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2	4 4 4 4 5 5 4 5 5 5 5 5 5 5 5 5 5 5 5 5		2

Fig. 4d

ENVELOPE REGION (4/5)

GENOTYPE	nedenborene krodenningen begrocken bereit bester be	181 CGCGAAAGGC	181 CGCGAAAGGC	181 CGCGAAAGGC	181 CGCGAAAGGC	181 CGCGAAAGGC		GII 181 CGCGAAAGGC	181		181 CGCGAAAGGC	181 CGCGAAAGGC			计时间 计计算 医多种 医多种 医多种性 医多种性 医多种性 医多种性 医多种性 医多种性 医	GIII 181 TGCGAAA	181	01645666786666846666690476666666888888888888888888888888888888	GIV 181 CGCGAAA	181 CGCGAAA
	GC CTTGTGGTAC	GC CTTGTGGTAC	GC CTTGTGGTAC	GC CTTGTGGTAC	GC CITGIGGIAC	GC CTTGTGGTAC	;0000000000000000000000000000000000000	GC CITGIGGIAC		GC CTTGTGGTAC	SGC CTTGTGGTAC	SGC CITGIGGIAC		SGC CITGIGGIAC		TGCGAAAGGC CITGIGGIAC TGCCTGAIAG GGTGCTTGCG AGTGCCCCGG GAGGTCTCGT	tocgaaagge citotoggiae toccidatag gotoctigeg agigeeeegg gaggietegi		COCGAAAGGC CIIGIGGIAC IGCCIGAIAG GGIGCFIGCG AGIGCCCCGG GAGGICICGI	COCGAAAGGC CTIGTGGTAC TGCCTGATAG GGTGCTTGCG AGTGCCCCGG GAGGTCTCGT
•	TOCCTOATAG		TGCCTGATAG	TGCCTGATAG		TGCCTGATAG		TGCCTGATAG	TGCCTGATAG	TGCCTGATAG	TGCCTGATAG	TGCCTGATAG	TGCCTGATAG	TGCCTGATAG		TOCCTGATAG	TGCCTGATAG		TGCCTGATAG	TGCCTGATAG
	TOCCIOATAO GOTOCITOCO	TOCCTOATAG GGIGCITGCG AGTGCCCCGG GAGGICTCGI	TGCCTGATAG GGTGCTTGCG	TGCCTGATAG GGTGCTTGCG	TOCCEGATAG GOTGCTTGCG AGTGCCCCGG	CITGIGGIAC IGCCIGATAG GGIGCITGCG AGIGCCCGG		CITGIGGIAC IGCCIGATAG GGIGCIIGCG	CTTGTGGTAC TGCCTGATAG GGTGCTTGCG	GOTGCTTGCG	GGTGCTTGCG	GGTGCTTGCG	CTTGTGGTAC TGCCTGATAG GGTGCTTGCG	TOCCIGATAG GGTGCTTGCG AGTGCCCCGG	1.以外的日日日本新年的	GGTGCTTGCG	ggreerrece		GGTGCTTGCG	GGTGCTTGCG
	AGTGCCCCGG GAGGTCTCGT	AGTOCCCCGG	AGTGCCCCGG	AGTGCCCCGG	AGTGCCCCGG	AGTGCCCCGG		AGTGCCCCGG	AGTGCCCCGG	AGTGCCCCGG	AGTGCCCCGG	AGTGCCCCGG	AGIGCCCGG	AGTGCCCCGG		AGTGCCCGG	AGTCCCCCGG	30 24 5 4 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5	AGTOCCCCGG	AGTGCCCCGG
	gaggtetegt	GAGGTCTCGT	GAGGICTCGT	GAGGTCTCGT	GAGGICICGI	GAGGTCTCGT		GAGGICTCOT	GAGGTCTCGT	GAGGTCTCGT	GAGGICICGI	GAGGICICGI	GAGGICICGI	GAGGTCTCGT		GAGGICICGI	GAGGTCTCGT		GAGGICICGI	GAGGICICGI

252 Total

12/2/

F'19. 40 s'UT Region (5/5)

2			
ID NUMBER	GENOTYPE		
			1900年末年日日第四日日日日日日日日日日日日日日日日日日日日日日日日日日日日日日日日日
33	ΙĐ	241	AGACCGIGCA CC
34		241	AGACCGIGCA CC
35		241	AGACCGIGCA CC
36		241	AGACCGIGCA CC
37		241	AGACCGIGCA CC
38		241	AGACCGIGCA CC
39	.110	241	AGACCGIGCA CC
40		241	AGACCGTGCA TC
41		241	AGACCGIGCA CC
42		241	AGACCGTGCA CC
43		241	AGACCGIGCA CC
44		241	AGACCGIGCA CC
45		241	AGACCGIGCA CC
4	1116	241	AGACCGTGCA TC
47		241	AGACCGICCA IC
	11	13	D
8	QIA	241	AGACCGTGCA AC
49		241	AGACCGIGCA AC

Fig. 5a

pressionered de contra de

) 5 #
ATGAGCACGA ATCCTAAACC TCAAAAAAA AACAAACGTA ACACCAACCG TCGCCCACAG ATGAGCACGA ATCCTAAACC TCAAAGAAA ACCAAACGTA ACACCAACCG TCGCCCACAG ATGAGCACGA ATCCTAAACC TCAAAGAAA ACCAAACGTA ACACCAACCG TCGCCCACAG ATGAGCACGA ATCCTAAACC TCAAAGAAA ACCAAACGTA ACACCAACCG TCGCCCACAG ATGAGCACGA ATCCTAAACC TCAAAGAAAA ACCAAACGTA ACACCAACCG TCGCCCACAG	ATGAGCACGA ATCCTAAACC TCAAAGAAAA ACCAAACGTA ACACCAACCG CCGCCCACAG ATGAGCACAA ATCCTAAACC TCAAAGAAAA ACCAAACGTA ACACCAACCG CCGCCCACAG ATGAGCACAA ATCCTAAACC CCAAAGAAAA ACCAAACGTA ACACCAAACG TCGCCCACAG ATGAGCACGA ATCCTAAACC TCAAAGAAAA ACCAAACGTA ACACCAACGG CCGCCCACAG ATGAGCACGA ATCCTAAAC TCAAAGAAAA ACCAAACGTA ACACCAACCG CCGCCCACAG ATGAGCACGA ATCCTAAACC TCAAAGAAAA ACCAAACGTA ACACCAACCG CCGCCCACAG ATGATAACC TCAAAGAAAA ACCAAACGTA ACACCAACCG CCGCCCACAG ATGAGCACGC TCAAAGAAAA ACCAAACGTA ACACCAACCG CCGCCCACAG ATGAGCAAACC TCAAAGAAAA ACCAAACGTA ACACCAACCG CCGCCCACAG ATGAGCACCA CCAAACGAACGTA ACACCAACCG CCGCCCACAG	ATGAGCACAA ATCCTAAACC TCAAAGAAA ACCAAAAGAA ACACTAACCG CCGCCCACAG ATGAGCACAA ATCGTCAACC TCAAAGAAAA ACCAAAAGAA ACACTAACCG CCGCCCACAG
ACACCAACCG ACACCAACCG ACACCAACCG ACACCAACCG ACACCAACCG	ACACCAACCG ACACCAACCG ACACCAACCG ACACCAACCG ACACCAACCG ACACCAACCG ACACCAACCG ACACCAACCG	ACACTAACCG ACACTAACCG
TCAAAAAAA ACAAACGTA TCAAAGAAA ACCAACGTA TCAAAGAAA ACCAACGTA TCAAAGAAA ACCAAACGTA TCAAAGAAA ACCAAACGTA TCAAAGAAA ACCAAACGTA	ACCAAACGTA ACCAAACGTA ACCAAACGTA ACCAAACGTA ACCAAACGTA ACCAAACGTA	ACCAAAAGAA ACCAAAAGAA
ATGAGCACGA ATCCTAAACC TCAAAAAAA AACAAACGTA ACACCAACCG ATGAGCACGA ATCCTAAACC TCAAAGAAAA ACCAAACGTA ACACCAACCG ATGAGCACGA ATCCTAAACC TCAAAGAAAA ACCAAACGTA ACACCAACCG ATGAGCACGA ATCCTAAACC TCAAAGAAAA ACCAAACGTA ACACCAACCG ATGAGCACGA ATCCTAAACC TCAAAGAAAA ACCAAACGTA ACACCAACCG	TCAAAGAAA TCAAAGAAA CCAAAGAAA TCAAAGAAA TCAAAGAAA TCAAAGAAA	TCAAAGAAA TCAAAGAAAA
ATCCTAAACC ATCCTAAACC ATCCTAAACC ATCCTAAACC ATCCTAAACC	ATCCTAAACC ATCCTAAACC ATCCTAAACC ATCCTAAACC ATCCTAAACC ATCCTAAACC	ATCCTAAACC ATCCTCAACC
52 DI ATGAGCACGA ATCCTAAACC TCAAAAAAA AACAACGTA ACACCAACGG TCGCCCACAG 53 1 ATGAGCACGA ATCCTAAACC TCAAAGAAAA ACCAAACGTA ACACCAACCG TCGCCCACAG 54 1 ATGAGCACGA ATCCTAAACC TCAAAGAAAA ACCAAACGTA ACACCAACGG TCGCCCACAG 55 1 ATGAGCACGA ATCCTAAACC TCAAAGAAAA ACCAAACGTA ACACCAACCG TCGCCCACAG 57 1 ATGAGCACGA ATCCTAAACC TCAAAGAAAA ACCAAACGTA ACACCAACCG TCGCCCACAG	58 GII 1 ATGAGCACGA ATCCTAAACC TCAAAGAAA ACCAAACGTA ACACCAACCG CCGCCACAG 59 1 ATGAGCACAA ATCCTAAACC TCAAAGAAA ACCAAACGTA ACACCAACCG CCGCCACAG 60 1 ATGAGCACAA ATCCTAAACC CCAAAGAAA ACCAAACGTA ACACCAACCG TCGCCCACAG 61 1 ATGAGCACGA ATCCTAAACC TCAAAGAAAA ACCAAACGTA ACACCAACCG CCGCCCACAG 62 1 ATGAGCACGA ATCCTAAACC TCAAAGAAAA ACCAAACGTA ACACCAACCG CCGCCCACAG 63 1 ATGAGCACGA ATCCTAAAGAAAA ACCAAACGTA ACACCAACCG CCGCCCACAG 64 1 ATGAGCACGA ATCCTAAAGAAAA ACCAAACGTA ACACCAACCG CCGCCCACAG 65 1 ATGAGCACGA ATCCTAAAGAAAA ACCAAACGTA ACACCAACCG CCGCCCACAG	65 GIII 1 ATGAGCACAA ATCCTAAACC TCAAAGAAAA ACCAAAAGAA ACACTAACCG CCGCCCACAG 66 1 ATGAGCACAA ATCGTCAACC TCAAAGAAAA ACCAAAAGAA ACCAAAAGCG CCGCCCACAG
	нннннн	
01	115	IIIO .
សល់ក្រស សម្មភាស ស្រួងស្រួស	10 10 00 00 00 00 00 00 00 00 00 00 00 0	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

Fig. 5b

CORE REGION (2/9)

SEQUENCE ID NUMBER	GENOTYPE					•		
2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2		61	nsennannnannnannnannnannnannnannnannnann	TCCCGGGTGG	usessessessessessessessessessessessesses	GITGGTGGAG	TITACTIGIT	GCCGCGCAGG
i in	}	61	GACGICAAGT	TCCCGGGTGG	CGGTCAGATC	GTIGGIGGAG	TITACITGII	GCCGCGCAGG
N 4	•	61	GACGITAAGT	TCCCGGGTGG	CGOTCAGATC	GTTGGTGGAG	TITACITGIT	GCCGCGCAGG
22		61	GACGTCAAGT	TCCCGGGTGG	CGGTCAGATC	GTTGGTGGAG	TITACITIGIT	GCCGCGCAGG
26		61	GACGICAAGI	TCCCGGGTGG	CGGTCAGATC	GIIGGIGGAG	TITACITECI	GCCGCGCAGG
5,7		19	GACGICAAGI	TCCCGGGTGG	CGGTCAGATC	GTTGGTGGAG	TCCCGGGTGG CGGTCAGAIC GITGGTGGAG TITACTIGIT GCCGCGCAGG	GCCGCGCAGG
0 14 15 15 15 15 15 15 15 15 15 15 15 15 15		# n n n						65.88.88.88.88.88.88.88.88.88.88.88.88.88
က	119	70	TOWNINGS	ラランクラランンフェ	TOWNING TOWNING		מדומנות האושרינות הייתימנים	204727272
8		61	GACGICAAGI	TCCCGGGCGG	TGGTCAGATC	GTIGGIGGAG	TITACCIGIL	CCCCCCAGG
9	•	61	GACGTCAAGT	TCCCGGGCGG	TGGTCAGATC	GILGGIGGAG	TITACCIGII	GCCGCGCAGG
61		61	GACGICAAGI	TCCCGGGCGG	TGGTCAGATC	GTTGGTGGAG	TITACITGII	GCCGCGCAGG
62	•	19	GACGTCAAGT	TCCCGGGCGG	TGGTCAGATC	GTTGGTGGAG	TITACCIGIT	GCCGCGCAGG
63		19	GACGICAAGI	TCCCGGGCGG	TGGTCAGATC	GITGGTGGAG	TITACITICAL	GCCCCCAGG
64		19	GACGICAAGI	GACGICAAGI ICCCGGGCGG	TOGICAGAIC	GTTGGTGGAG	TITACCIGIT	GCCGCGCAGG
	100000000000000000000000000000000000000	# F E	GACGTCAAGT	TCCCGGGCGG	BEST TEGECAGATE	GTTGGCGGAG	TATACTIGCT	countrocates communicate control of the control of
) v	;	5 6	GACGTCAAGT	TCCCGGGCGG	TGGTCAGATC	GITGGCGGAG	TATACTIGIT	CALCATORAGE TECTORGETC TOTTO OF TOTOGETCAGA TATACTIVET GEOGEGAG
>		:						

Fig. 50

CORE REGION (3/9)

52 01 121 GGCCTAGAT TGGGTGTGC CGCGACGAGA AAGACTTCCG AGCGGTCGCA ACCTCGAGGT 53 121 GGCCCTAGAT TGGGTGTGCG CGCGACGAGG AAGACTTCCG AGCGGTCGCA ACCTCGAGGT 54 121 GGCCCTAGAT TGGGTGTGCG CGCGACGAGG AAGACTTCCG AGCGGTCGCA ACCTCGAGGT 55 121 GGCCCTAGAT TGGGTGTGCG CACGACGAGG AAGACTTCCG AGCGGTCGCA ACCTCGAGGT 56 121 GGCCCTAGAT TGGGTGTGCG CGCGACGAGG AAGACTTCCG AGCGTCGCA ACCTCGAGGT 57 121 GGCCCTAGAT TGGGTGTGCG CGCGACGAGG AAGACTTCCG AGCGTCGCA ACCTCGAGGT 57 121 GGCCCTAGAT TGGGTGTGCG CGCGACGAGG AAGACTTCCG AGCGTCGCA ACCTCGAGGT							
	121	GCCCTAGAT	BOBECTAGAT TOOTOTOGC COCCACGAG AAGACTICC AGGGTCGC ACCTCAGGG	CGCGACGAGA	AAGACTTCCG	AGCGGTCGCA	ACCTCGAGGT
	121	GGCCCTAGAT	IGGGIGIGG CGCGACGAGG AAGACTICCG	CGCGACGAGG	AAGACTTCCG	AGCGGTCGCA ACCTCGAGGT	ACCTCGAGGT
55 53 54	121	GGCCCTAGAT	recorded cecarcage	CGCGACGAGG	AAGACTTCCG	AGCGGTCGCA	ACCICGAGGI
55 57	121	GOCCCTAGAT	TOGGTGTGCG CACGACGAGG	CACGACGAGG	AAGACITCCG	AGCGGTCGCA	ACCTCGAGGT
57	121	GGCCCTAGAT	TOOUTGIEGG COCCACCAGG AAGACTICCG	CGCGACGAGG	AAGACTTCCG	AGCGGTCGCA	ACCTCGAGGT
	121	GGCCCTAGAT	receretece	CGCGACGAGG	AAGACTTCCG	TOGGIGIGG CGCGACGAGG AAGACTICCG AGCGGICGCA ACCICGIGGI	ACCICGIGGI
58 GII	121	GGCCCCAGGT	GOCCCCAGOI IGGOIDIGCO COCOACIAGO AAGACIICCO AGCGGICGCA ACCICGIAGA	りりはていなりいりつ	AAGACTICCG	AGCGGTCGCA	ACCICATORA
53	121	GGCCCCAGGT		CGCGACTAGG	AAGACTICCO	TOGGIGICO CCCGACIAGO AAGACIICCO AGCGGICOCA ACCICGIGGA	ACCTCGTGGA
09	121	GGCCCCAGGT	_	CGCGACTAGG	AAGACTTCCG	IGGGIGIGG CGCGACIAGG AAGACIICCG AGCGGICGCA ACCICGIGGA	ACCTCGTGGA
19	121	GCCCCAGGT	regererece	CGCGACTAGG	AAGACTTCCG	AGCGGTCGCA ACCTCGTGGA	ACCTCGTGGA
62	121	GGCCCCAGGT		CGCGACTAGG	AAGACTTCCG	TEGETETECE CECEACTAGE AAGACTICCE ACCECECA ACCICGIGA	ACCTCGTGGA
63	121	GGCCCCAGGT		CGCGACTAGG	AAGACTTCCG	TEGETETECE CECEACIAGE AAGACTICCE AGCGGICGCA ACCICGIGEA	ACCTCGTGGA
64	121	GGCCCCAGGT	Tegererece	CGCGACTAGG	AAGACTTCCO	TOGITGIGG CCCGACTAGG AAGACTICCG AGCGGTCGCA ACCICGIGGA	ACCTCGTGGA
				1000000000000			
IIID 65	I 121	GGCCCGAGAT	GGCCCGAGAT TGGGTGTGCG CGCGACGAGG AAAACTTCCG AACGATCCCA GCCACGCGGA	CGCGACGAGG	AAAACTTCCG	AACGATCCCA	GCCACGCGGA
66	121	GGCCCCAGGT	goccccaggi iggalataca coccaccaaga aaaaciiicca aacgoiicca occacciiggo	CCCCACGAGG	AAAACTTCCG	AACGGTCCCA	GCCACGIGGG

Fig. 5d

CORE REGION (4/9)

	10 11 10 10 10 10 10 10 10 10 10 10 10 1			A B B B B B B B B B B B B B B B B B B B	00000000000000000000000000000000000000			40000000000000000000000000000000000000
SECUENCE	•					-		
ID NUMBER	GENOTYPE							
52	GI	181	AGACGTCAGC	CTATCCCCAA	GGCTCGTCGG	AGACGICAGC CIAICCCCAA GGCICGICGG CCCGAGGGCA GGACCIGGGC ICAGCCCGGG	GGACCTGGGC	ICAGCCCGGG
i C	1	181	AGACGICAGC	CTATCCCCAA	GGCGCGTCGG	ADACGICAGC CIAICCCCAA GGCGCGICGG CCCDAGGGCA GGACCIGGGC ICAGCCCGGG	GOACCIOGGC	TCAGCCCGGG
3 2		181	AGACGTCAGC	CTATCCCTAA	GGCGCGTCGG	APAPETER CTATCCCTAA GGCGCGTCGG CCCGAGGGCA	GGACCTGGGC	TCAGCCCGGG
ð		707	000000000000000000000000000000000000000		大力をして出ている	べしかがれないが、 木が生むが出りが、 木がりがはませる	CONTRACTOR OF	からいしいからい
ស		181	AGACGICAGC	CCATCCCCAA	MOTTOTTON	CCCOROCCA	200012200	
56		181	AGACGTCAGC	CTATCCCCAA	GGCACGTCGG	CTAICCCCAA GGCACGICGG CCCGAGGGIA	DEACCTORGE TCAGCCCGGG	TCAGCCCGGG
57		181	AGACGCCAGC	CTATCCCCAA	ggcgcgrcgg	adacgecage ctateceena ggegegtegg eeegagggea ggaeetggge teageeegg	GGACCTGGGC	TCAGCCCGGG
S. C.	611	181	AGGCGACAAC	CIATCCCCAA	GGCTCGCCAG	AGGCGACAAC CTATCCCCAA GGCTCGCCAG CCCGAGGBCA GGGCCTGGGC TCAGCCCGGG	GGGCCTGGGC	TCAGCCCGGG
9 6		181	AGGCGACAAC	CTATCCCCAA	GGCTCGCCAG	AGGCGACAAC CIPICCCCAA GGCICGCCAG CCCGAGGGCA GGGCCIGGGC ICAGCCCGGG	GGGCCTGGGC	TCAGCCCGGG
) Y	•	181	AGGCGACAAC	CTATCCCCAA	GGCTCGCCGG	AGGCGACAAC CTATCCCCAA GGCTCGCCGG CCCGAGGGCA GGTCCTGGGC	GGTCCTGGGC	TCAGCCCGGG
3 5		181	AGGCGACAAC	CTATCCCCAA	GGCTCGCCAG	ARREGACIAR CTATCCCCAA GGCTCGCCAG CCCGAGGGTA GGGCCTGGGC	GGGCCTGGGC	TCAGCCCGGG
4 6		181	AGGCGACAAC	CTATCCCCAA	GGCTCGCCGG	CTATCCCCAA GGCTCGCCGG CCCGAGGGCA GGGCCTGGGC	GGGCCTGGGC	TCAGCCCGGG
		181	AGGCGACAAC	CTATCCCCAA	GGCTCGCCGG	CIATCCCCAA GGCTCGCCGG CCCGAGGGCA GGGCCTGGGC TCAGCCCGGG	CCCCTCCCC	TCAGCCCGGG
3 ;			してくじんじょうじゃ	ママンシンサイサン	CCCACCAC	ACCESSANCE FEMALLICATION GREENCHICA COCCANGGGAN GGGCCTGGGC TCAGCCCGGG	CCCCTCCC	TCAGCCCGGG
\$	•	101	つませいよういううちょ					
# # # # # # # # # # # # # # # # # # #			aeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeee	CCATCCCTAA	AGATCGTCGC	BERNHERERE ESERTE EN PROCECTAR AGAICATCO ACCOCTOBCA AGICCTOGGG AAGGCCAGGA	AGTCCTGGGG	**************************************
3 1	;		して ないしいしつ でん	A STORY OF	AGATCGGCGC	**************************************	AGTCCTGGGG	GAAGCCAGGA
6		101	101111111111111111111111111111111111111					

Fig. 5e

CORE REGION (5/9)

Enumeraciones estatos estatos

25	19	241	TACCCTTGGC	CCCTCTATEG	CCCTCTATEG CAATGAGGC TGCGGGTGGG CGGGATGGCT	roccoorage	CGGGATGGCT	52 di 241 TACCCTTGGC CCCTCTATGG CAATGAGGGC TGCGGGTGGG CGGGATGGCT CCTGTCTCCC
53.		241	TACCCTTGGC		CCCTCTATEG CAATGAGGET TGCGGGTGGG CGGGATGGCT	TGCGGGTGGG	CGGGATGGCT	CCIGICICC
54		241	TACCCCTGGC	CCCTCTATGG	TAATGAGGGT	TGCGGATGGG	CGGGATGGCT	CCTGTCCCC
55		141	TACCCTTGGC		CCCTCTATGG CAATGAGGGC	TGCGGGTGGG	CGGGATGGCT	cereterece
56		241	TACCCTTGGC		CCCTCTATGG CAATGAGGGT	TGCGGGTGGG	CGGGATGGCT	cerateree
51		241	TACCCITGGC	CCCTCTATGG	CAATGAGGGT	rccccorcca	CCCTCTATGG CAATGAGGGT TGCGGGTGGG CGGGATGGCT	CCTGICICCC
		1000000						
56	110	241	TACCCTTGGC	CCCTCTATGG	CAATGAGGGT	ATGGGGTGGG	TACCCTTGGC CCCTCTATGG CAATGAGGGT ATGGGGTGGG CAGGATGGCT CCTGTCACCC	CCIGICACCC
59		241	TACCCTTGGC	CCCTCTATGG	CAACGAGGGT	ATGGGGTGGG	TACCCITIGGC CCCTCTATGG CAACGAGGGT AIGGGGTGGG CAGGATGGCT	CCTGTCACCC
9	•	241	TACCCTIGGC		CCCTCTATGG CAACGAGGT ATGGGGTGGG CAGGATGGCT	ATGGGGTGGG	CAGGATGGCT	CCTGTCACCC
61		241	TACCCTTGGC	CCCTCTATGG	CCCTCTATGG CAATGAGGGT ATGGGGTGGG CAGGGTGGCT	ATGGGGTGGG	CAGGOTGGCT	CCIGICCCC
62		341	TATCCTTGGC		CAATGAGGGT	CIGGGGIGGG	CCCTCTATES CAATGAGGGT CTGGGGTGGG CAGGATGGCT	CCTGTCACCC
63		241	TACCCTTGGC		CAATGAGGGT	ATGGGGTGGG	CCCTCTATGG CAATGAGGGT ATGGGGTGGG CAGGATGGCT CCTGTCACCC	CCTGTCACCC
99		241	TACCCCTGGC	CCCTCTATGG	CAATGAGGGT	ATGGGGTGGG	IACCCCTGGC CCCTCTATGG CAATGAGGT ATGGGGTGGG CAGGATGGCT CCTGTCACCC	CCTGICACCC
								计计计址形式设计记录 计记录记录 医多种多种 医多种多种 医多种多种 医多种多种 医多种 医多种 医多种 医多种
92	IIID	241	TATCCTTGGC	CCCTGTATGG	GAATGAGGGT	CICGGCIGGG	TATCCTTGGC CCCTGTATGG GAATGAGGGT CTCGGCTGGG CAGGGTGGCT CCTGTCCCCC	ccrorecee
99		241	TACCCTTGGC	CCCTGTAIGG	TACCTIGGC CCCIGIAIGG GAAIGAGGI CICGGCIGGG CAGGGIGGCI CCIGIACCC	CTCGGCTGGG	CAGGGTGGCT	CCTGTCCCC

18/2/

Fig. 5

CORE REGION (6/9)

VACUE OF 01	GENOTYPE							
E3020222 52	cadestrassacoosossas 52 GI	301	**************************************	GGCCTAGCTG	######################################		normentonenonungenannungensensensensensensonsonungensensensensensensensensensensensensense	CAATITGGG
53		301	CGTCCCTCTC	GGCCTAGTTG	GGGCCCCACA	GACCCCCGGC	GECCTAGITIC GESCECEACA GACCECESSE STABSICES CAATITISSEI	CAATTIGGG
		301	CORGCICIC	GGCCTAGTTG	GGGCCCTACA	GACCCCCGGC	GGCCTAGTIG GGGCCCTACA GACCCCCGGC GTAGGICGCG CAAITIGGGI	CAATTTGGG
. R.		301	COTCCTCTC	GGCCTAGCTG	GGGCCCCACA	GACCCCCGGC	GOCCTAGCTO GOGCCCCACA GACCCCCGGC GTAGGTCGCG CAATTTGGG1	CAATTTGGGJ
55.		301	CGCGGCICIC	GGCCTAACTG	GGGCCCCACA	GACCCCCGGC	OGCCTAACTG GGGCCCCACA GACCCCGGC GTAGGTCGCG CAATTTGGGT	CAATTTGGG
57		301	CGIGGCICIC	GGCCTAGCTG	GGGCCCCACA	GACCCCCGGC	condecitic deceivates edecetaca daceeecoo diadorede caaiiises	CAATITGGG
		X 11 12 15 15 15 15 15 15 15 15 15 15 15 15 15						
58	110	301	CGIGGCICIC	GGCCTAGITG	GGGCCCCACG	GACCCCGGC	GGCCTAGITG GGGCCCCACG GACCCCGGC GIAGGICGCG IAATIIGGG	TAATTTGGG
59		301	CGIGGCICIC	GGCCTAGITG	GGGCCCCACG	GACCCCGGC	GOCCIAGING GGGCCCCACG GACCCCCGGC GIAGGICGCG TAAINIGGG	TAATTTGGG
09	•	301	CGCGGCTCCC	GGCCTAGITG	GGGCCCCACG	GACCCCGGC	GECCTAGIIG GGGCCCCACG GACCCCCGGC GIAGGICGCG IAAIIIGGG	TAATTIGGG
61		301	cdceecaccc	GGCCTAGTTG	GGGCCCCACA	GACCCCCGGC	cácescroco seceradara assecceaca sacecesses stasseces raarrasses	TAATTIGGG
62		301	CGCGGCTCTC	GGCCTAGCTG	GGGCCCTACC	GACCCCCGGC	GOCCIAGCIO GOGCCCIACO GACCCCGGC GIAGGICGCG CAACIIGGG	CAACTTGGG
63		301	COTOGITICIC	GGCCTAGITG	GGGCCCCACG	BACCCCGGC	congornere decerating addeceased aacceeege gradgreed	CAATTIGGG
64		301	CGCGGCTCCC	GGCCIAGITIG	GGGCCCCAAA	GACCCCCGGC	CGCGGCICCC GGCCIAGTIG GGGCCCCAAA GACCCCCGGC GIAGGICGCG IAAIIIGGG	TAATTIGGG
		11 11 11	1000时以外以外以外的企业,不是一个一个一个一个一个一个一个一个一个一个一个一个一个一个一个一个一个一个一个					
65	0111	301	COTGGCTCTC	GCCCTTCATG	GGGCCCCACT	GACCCCCGGC	cotogetete geetiteato oggeceeact gaceeegge atagategeg caactiegg	CAACTIGGG
99		301	CGCGGTTCTC	GCCCTTCATG	GGGCCCCACT	GACCCCCGGC	cacattata accetteata aggececact gaceeeegge atagateaeg caacttagg	CAACTIGGG

Fig. 5c

CORE REGION (7/9)

SEQUENCE GENOTYPE

		965555			AGGETHEBEHEBEHEBEHEREBEHEREBEHERGTOOFGETEREFERHEBEHEBEHEBEHEBEHEBEHEBEHEBEHEBEHEBEHE	GCCGACCTCA	TGGGGTACAT	enspannennnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
70	;	1	774					
53		361	AAGGTCATCG	ATACCCTTAC	AAGGICAICG AIACCCIIAC GIGCGGCIIC GCCGACCACA	GCCGACCACA	TOGGGTACAT ACCGCTCGTC	ACCECTCGTC
54		361	AAGGTCATCG	ATACCCTCAC	AAGGICAICG AIACCCICAC GIGGGGCIIC GCCGACCACA	GCCGACCACA	TGGGGTACAT TCCGCTCGTT	rcccrccrr
55		361	AAGGTCATCG	ATACCCITAC	ATACCCITAC GIGCOGCIIC	GCCGACCTCA	TGGGGTACAT	ACCGCTCGTC
26		361	AAGGICAICG	ATACCCTTAC	ATACCCTIAC GIGCGCTIC	GCCGACCTCA	TGGGGTACAT	ACCGCTCGTC
57		361	AAGGTCATCG	ATACCCTTAC	AAGGICAICG AIACCTITAC GIGGGCIIC GCCGACCICA IGGGGIACAI ACCGCICGIC	GCCGACCTCA	TGGGGTACAT	ACCOCTOOTC
					26086868686			
80	119	361	AAGGTCATCG	ATACCCTCAC	AAGGICAICG AIACCCICAC AIGCGGCIIC GCCGACCICA IGGGGIACAI ICCGCICGIC	GCCGACCTCA	TGGGGTACAT	recerteare
8		361	AAGGTCATCG	ATACCCTCAC	AIACCCICAC AIGCGGCIIC GCCGACCICA IGGGGIACAI ICCGCIIGIC	GCCGACCICA	TGGGGTACAT	TCCGCTTGTC
9	•	361	AAGGICAICG		ATACCCICAC AIGCGGCTIC	GCCGACCTCA	TGGGGGTACAT	recerrenc
61		361	AAGGTCATCG		ATACCCICAC AIGCGGCIIC GCCGACCICA	GCCGACCTCA	TGGGGGTACAT	TCCCCTCGTC
62		361	AAGGTCATCG		ATACCCTTAC GIGGGGCTIC	GCCGACCTCA	TGGGGTACAT	TCCGCTCGTC
6.0		361	AAGATCATCG	ATACCCTCAC	ATACCCICAC GIGGGGCIIC	GCCGACCTCA	TGGGGTACAT	TCCGCTCGTC
4		361	AAGGTCATCG	ATACCCTCAC	AAGGICAICG AIACCTICAC AIGCGGCIIC GCCGACCICA	GCCGACCICA	TGGGGTACAT	rccerrcirc
. vo	GIII	361	AAGGTCATCG	ATACCCTAAC	AAGGICAICG AIACCCIAAC GIGCGGIIIT GCCGACCICA	GCCGACCTCA	TGGGGTACAT	TGGGGTACAT TCCCGTCATC
9	ı	361	AAGGTCATCG	ATACCCTAAC	GICIGCITII	GCCGACCTCA	TGGGGTACAT	AAGGTCATCG ATACCCTAAC GTGTGGTTTT GCCGACCTCA TGGGGTACAT TCCCGTCGGT
1 1 1								50个的分子与约约日的公司的特别的特别的现在分词的现在分词 医多种性皮肤 医多种种 医多种种皮肤 医多种皮肤 医多种皮肤 医多种种皮肤 医多种皮肤 医多种皮肤皮肤 医多种皮肤皮肤 医多种皮肤皮肤 医多种皮肤 医多种皮肤皮肤皮肤皮肤皮肤皮肤皮肤皮肤皮肤皮肤皮肤皮肤皮肤皮肤皮肤皮肤皮肤皮肤皮肤

20/21

5h	
Fig.	

CORE REGION (8/9)

SEQUENCE									
Decoration 10 10 10 10 10 10 10 1	0 0 0 0 0 0	10 10 10 10 10 10 10 10 10 10 10 10 10 1	4 0 0 0						
NUMBER GENOTYPE	SEQUENCE								
52 GI 421 GGCGCCCTC TTGGAGGGC TGCCAGGGC CTGGCGCATG GCGTCCGGGT TCTGGAAGA 53 421 GGCGCCCTC TTGGAGGCG TGCCAGGGC CTGGCGCATG GCGTCCGGGT TCTGGAAGA 54 421 GGCGCCCTC TTGGAGGCG TGCCAGGGC CTGGCGCATG GCGTCCGGGT TCTGGAAGA 55 421 GGCGCCCTC TTGGAGGCG TGCCAGGGC CTGGCGCATG GCGTCCGGGT TCTGGAAGA 56 421 GGCGCCCTC TTGGAGGCG TGCCAGGGC CTGGCGCATG GCGTCCGGGT TCTGGAAGA 58 GII 421 GGCGCCCTC TTGGAGGCG TGCCAGGGC CTGGCGCATG GCGTCCGGGT TCTGGAAGA 59 421 GGCGCCCTC TTGGAGGCG TGCCAGGGC CTGGCGCATG GCGTCCGGGT TCTGGAAGA 50 421 GGCGCCCCT TAGGGGGCG TGCCAGGGC CTGGCGCATG GCGTCCGGGT TCTGGAGGA 60 421 GGCGCCCCC TAGGGGGCG TGCCAGGGC CTGGCGCATG GCGTCCGGGT TCTGGAGGA 61 GGCGCCCCC TAGGGGGCG TGCCAGGGC CTGGCGCATG GTCCCGGGT TCTGGAGGA 62 421 GGCGCCCCC TAGGGGGCG TGCCAGGGC CTGGCGCATG GTCCCGGGT TCTGGAGGA 63 GGCCCCCC TAGGGGGCG TGCCAGGGC CTGGCGCATG GTCCCGGGT TCTGGAGGA 64 GGCGCCCCC TAGGGGGCG TGCCAGGGC CTGGCGCATG GTCCCGGGT TCTGGAGGA 64 GGCGCCCCC TAGGGGGCG TGCCAGGGC CTGGCGCATG GCGTCCGGGT TCTGGAGGA 64 GGCCCCCC TAGGGGGCC TGCCAGGGC CTGCCGCATG GCTCCCGGGT TCTGGAGGA TCTGCAGGCC TGCCAGGCC TGCCGCATG GCTCCCGGGT TCTGGAGGA TCTGCAGGCC TGCCAGGCC TGCCGCATG GCTCCCGGGT TCTGGAGGA TCTGCCAGGCC TGCCCCCATG GCTCCCGGGT TCTGCAGGC TTGCCAGGCC TTGCCAGGCC TTGCCCATG GCTCCCGGGT TCTGCAGGCC TTGCCCATG GCTCCCGGT TCTGCAGGC TCTCCCCATG GCTCCCGGT TCTGCAGGC TCTGCCCCATG GCTCCCGGT TCTGCAGGC TCTGCCCCATG GCTCCCGGT TCTGCAGGC TCGCCCCCT TAGGGGCC TGCCCCATG GCTCCCGGT TCTGCAGGC TCTCCCGGT TCTGCCCCCT TAGGGCC TGCCCCCT TAGGGCC TTGCCCCCT TCCCCCCT TCCCCCCT TCCCCCCT TCCCCCC	ID NUMBER	GENOTYPE							
53 421 GGCGCCCTC TTGGAGGCGC TGCCAGGGCT CTGGCCATG GCGTCCGGGT TCTGGAAGA			421		TTGGAGGCGC	TGCCAGGGC	CTGGCGCATG	GCGTCCGGGT	TCTGGAAGAC
54	53		421	CCCCCCTC		TGCCAGGGCT	CTGGCGCATG	GCGTCCGGGT	TCTGGAAGAC
10 10 10 10 10 10 10 10	. 54		421	OCCOCCUE	_	TGCCAGGGCC	CTGGCGCATG	GCGTCCGGGT	TCTGGAAGAC
10 10 10 10 10 10 10 10	52		421	GGCGCCCTC	TTGGAGGCGC	TGCCAGAGCC	CTGGCGCATG	GCGTCCGGGT	TCTGGAAGAC
57 421 GGGGCCCTC TTGGAGGGG TGCCAGGGCC CTGGCGCATG GCGTCCGGGT TCTGGAAGA	56		421	OGCCCCCTC		TGCCAGGGCC	CTGGCGCATG	GCGTCCGGGT	TCTGGAAGAC
1	24	-	421	GGCGCCCTC		TOCCAGGGCC	CTGGCGCATG	GCGTCCGGGT	TCTGGAAGAC
421 GGCGCCCCC TAGGGGCGC TGCCAGGGC CTGGCACATG GTGTCCGGGT TCTGGAGGA 60			421		TTAGGGGCGC	TGCCAGGGCC	TTGGCGCATG	GCGTCCGGGT	TCTGGAGGAC
60 421 GGGGCCCCC TAGGGGCGC TGCCAGGGCC CTGGCACATG GTGTCCGGGT TCTGGAGGA 61 421 GGCGCCCCC TAGGGGCGC TGCCAGGGC CTGGCGCATG GCGTCCGGGT TCTGGAGGA 62 421 GGCGCCCCC TAGGGGCGC TGCCAGGGC CTGGCGCATG GCGTCCGGGT TCTGGAGGA 63 421 GGCGCCCCC TAGGGGCGC TGCCAGGGC CTGGCGCATG GCGTCCGGGT TCTGGAGGA 64 421 GGCGCCCCC TAGGGGCGC TGCCAGGGC CTGGCGCATG GCGTCCGGGT TCTGGAGGA 65 GIII 421 GGCGCCCCC TAGGGGCGT TGCCAGGCC TGCCAGGGC TTGGCGAGGT TCTGGAGGA 66 GIII 421 GGCGCCCCG TTGGAGGCGT TGCCAAGGC CTCGCCCACG GAGTGAGGGT TCTGGAGGA 66 421 GGCGCCCCCG TTGGAGGCGT TGCCAAGCC CTCGCCCACG GAGTGAGGGT TCTGGAAGA	59		421	222222222	TAGGGGGCGC	TGCCAGGGCC	CTGGCACATG	GIGICCOGGI	TCTGGAGGAC
61 421 GGCGCCCCC TAGGGGCGC TGCCAGGGCC CTGGCGCATG GCGTCCGGGT TCTGGAGGA 62 421 GGCGCCCCC TAGGGGCGC TGCCAGGGC CTGGCGCATG GCGTCCGGGT TCTGGAGGA 63 421 GGCGCCCCC TAGGGGCGC TGCCAGGGC CTGGCGCATG GCGTCCGGGT TCTGGAGGA 64 421 GGCGCCCCT TAGGGGGCG TGCCAGGCC TGGCGCATG GCGTCCGGGT TCTGGAGGA 65 GIII 421 GGCGCCCCG TTGGAGGCGT TGCCAGAGCT CTCGCCCACG GAGTGAGGGT TCTGGAGGA 66 421 GGTGCCCCG TTGGAGGCGT TGCCAGAGCT CTCGCCCACG GAGTGAGGGT TCTGGAGGA	60	•	421	ეეეეეეენენე	TAGGGGGCGC	TGCCAGGGCC			TCTGGAGGAC
63 421 GGGGCCCCC TAAGGGGCG TGCCAGGGCC CTGGCGCATG GCGTCCGGGT TCTGGAGGAGA 63 421 GGCGCCCCC TAAGGGGCGC TGCCAGGGC CTGGCGCATG GCGTCCGGGT TCTGGAGGAGA 64 421 GGCGCCCCCT TAAGGGGGCG TGCCAGGGCC CTGGCGCATG GCGTCCGGGT TCTGGAGGAGGCG TGCCAGGGCC TGCCAGGGCC TGCCAGGGC TCTGGAGGT TCTGGAGGC TGCCAGGCC TGCCAGGCC TCTGGAGGT TCTGGAAGGT TCTGGAAGAGT CTGGCCAATG GGGTGAAGGT TCTGGAAGAGT TCTGGAAGGT TCTGCCAATG GGGTGAAGGT TCTGGAAGGT TCTGCCCAATG GGATGAAGGT TCTGGAAGGT TCTGGAAGGT TCTGGAAGGT TCTGGAAGGT TCTGCCCAATG GGATGAAGGT TCTGGAAGGT TCTGCCCAATG GGATGAAGGT TCTGGAAGGT TCTGCCCAATG GGATGAAGGT TCTGCCAATG GGATGAAGGT TCTGCCAATG GAATGAAGT TCTGCCAATG GAATGAAGAAGT TCTGCCAATG GAATGAAGAAGT TCTGCCAATG GAATGAAGAAGT TCTGCCAATG GAATGAAGAAGT TCTGCCAATG GAATGAAGAAGT TCTGCCAATG GAATGAAGAAGT TCTGCCAATGAAGAAGT TCTGCCAATG GAATGAAGAAGT TCTGCCAATG GAATGAAGAAGAAGAAGT TCTGCCAATG GAATGAAGAAGAA	61		421	00,000,000	TAGGGGGCGC	TGCCAGGGCC	CIGGCGCAIG		TCTGGAGGAC
63 421 GGGGCCCCCC TAGGGGGCGC TGCCAGGGCC CTGGCGCATG GCGTCCGGGT TCTGGAGGA 64 421 GGGGCCCCCT TAGGGGGCGC TGCCAGGGC CTGGCGCATG GCGTCCGGGT TCTGCAGGA 65 GII 421 GGGGCCCCCG TTGGAGGCGT TGCCAGAGCT CTCGCCCACG GAGTGAGGGT TCTGGAGGA 66 421 GGTGCCCCCG TTGGAGGCGT TGCCAGAGCT CTCGCCCATG GGGTGAGGGT TCTGGAAGA	62		421	2222228288	TTAGGGGCGC	TGCCAGGGCC	CTGGCGCATG	GCGTCCGGGT	TCTGGAGGAC
64 421 GGGGCCCCCT TAGGGGGGGC TGCCAGGGCC CTGGCGCATG GCGTCCGGT TCTGCAGGA 65 GIII 421 GGCGCCCCC TTGGAGGCGT TGCCAGAGCT CTCGCCCACG GAGTGAGGGT TCTGGAGG 66 431 GGTGCCCCC TTGGAGGTT CGCCAGAGCC CTTGCCCATG GGGTGAGGGT TCTGGAAG	63		421	222222222	TAGGGGGGG	TGCCAGGGCC	CTGGCGCATG	GCGICCGGGI	TCTGGAGGAC
	64		421	ECCCCCC	TAGGGGGGGG	TGCCAGGGCC	CTGGCGCATG	GCGICCGGGI	TCTGGAGGAC
GIII 421 421	11 11 11 11 11 11 11 11 11 11 11 11 11					11 12 12 12 12 12 12 12 12 12 12 12 12 1		11 11 11 11 11 11 11 11 11	
421	65	1119	421	<u> </u>	TTGGAGGCGT	TGCCAGAGCT	CTCGCCCACG	GAGTGAGGGT	TCTGGAGGAT
	99		421	ggracecea	rrdefector	CGCCAGAGCC	CITGCCCAIG	GGGTGAGGGT	TCTGGAAGAC

21/21

Fig. 51

CORE REGION (9/9)

					(A)A). MOTOGN GNOO	(6/6).			
11 11 11 11 11	8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8	0 0 0 0	# 11 11 11 11 11 11 11 11 11 11 11 11 11						# 11 12 13 13 14 17
SEQUENCE								•	
ID NUMBER GENO	GENOTYPE								
				************					11 11 11 11 11 11 11
52	19	481	GGCGTGAACT	ATGCAACAGG	GGCGTGAACT ATGCAACAGG GAACCTTCCT GGTTGCTCTT	GGTTGCTCTT	TCTCTATCTT	CCTTCTGGCC CTGCTCTCT	CIGCICICI
53		481	GGCGTGAACT	ATGCAACAGG	GAACCITCCI	GGTIGCTCTT	TCTCTATCTT	GGCGTGAACT AIGCAACAGG GAACCIICCI GGIIGCICIT ICICIAICIT CCIIICIGGCC CIGCICICI	CIGCICICI
.54		481	GGCGTGAACT	ATGCAACAGG	ATGCAACAGG GAATCTTCCT GGTTGCTCTT	GGIIGCICII	TCTCTATCTT	CCTTCTGGCC	CTICICICI
55		481	GGCGTGAACT	ATGCAACAGG	GGCGTGAACT ATGCAACAGG GAACCTTCCC GGTTGCTCTT	GGTTGCTCTT		CCTTCTGGCC	CTGCTCTCT
26		481	GGCGTGAACT	ATGCAACAGG	GGCGTGAACT ATGCAACAGG GAACCTTCCT GGTTGCTCTT	GGTTGCTCTT		CCTTCTGGCC	CIGCICICI
57		481	GGCGTGAACT	ATGCAACAGG	GAACCTTCCT	GGTTGCTCTT			CIGCICICI
									10 40 40 40 41 41 41
28	119	481	GGCGTGAACT	ACGCAACAGG	GAATCTGCCC	OGIIGCICCI	TITCIAICII	ACGCAACAGG GAATCTGCCC GGTTGCTCCT TTTCTATCTT CCTCTTGGCT CTGCTGTCC	CTGCTGTCC
59		481	GGCGTGAACT	ATGCAACAGG	GAATTTGCCC	GGTTGCTCTT	TCTCTATCTT	GGCGTGAACT AIGCAACAGG GAATTIGCCC GGIIGCICIT ICICIAICIT CCICIIGGCI CIGCIGICC	CIGCIGICC
9	•	481	GGCGTGAACT	ATGCAACAGG	GGCGTGAACT AIGCAACAGG GAATITGCCT	GGTTGCTCTT	GGTTGCTCTT TCTCTATCTT	CCICITIGGCI CIGCIGICC	CIGCIGICC
61		481	GGCGTGAACT	ATGCAACAGG	GGCGTGAACT ATGCAACAGG GAATCTGCCC	GGTTGCTCTT	TCTCTATCTT		Trecrerce
62	٠	481	GGCGTGAACT	ATGCAACAGG	GGCGTGAACT ATGCAACAGG GAATTTGCCC GGTTGCTCTT	GGTTGCTCTT			TIGCIGICC
63		481	GGCGTGAACT	ATGCAACAGG	GAATCTGCCC	GOTTGCTCCT	GGCGIGAACT AIGCAACAGG GAATCIGCCC GGINGCICCI TITCIAICII		TECTOTOC
64		481	GGCGTGAACT	ATGCAACAGG	GAATCIACCC	GOTTGCICIT	TCTCTATCTT		TIGCIGICC
20 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0		481	immentannessanssansannannannannannannannannannanna	ATGCAACAGG	GAATTTGCCC	GGTTGCTCTT	TCTCTATCTT	ensennensennnnnnnnnnnnnnnnnnnnnnnnnnnn	Sananann:
99		481	GGGATAAATT	GGGATAAATT ATGCAACAGG GAATCTGCCC	GAATCTGCCC				
	- 75		5月11日日日日日日日日日日日日日日日日日日日日日日日日日日日日日日日日日日日						

549 Total